IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: Jack et al. EXAMINER: Hutson

SERIAL NO.: 10/089,027 GROUP: 1652

FILING DATE: March 26, 2002

FOR: Incorporation of Modified Nucleotides by Archaeon DNA Polymerases and Related Methods

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Certificate of Mailing

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Mail Stop AF, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on May 4, 2006.

Leslie Goldberg

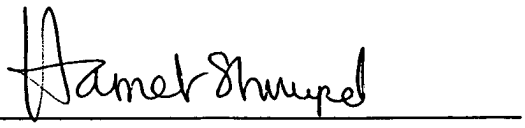
Sir:

Deletion of Inventors Pursuant to 37 C.F.R. §1.48(b)

Applicants request that inventors Philip Buzby and James DiMeo whose names appears on the declaration and power of attorney filed in the subject case, a copy of which is enclosed, be deleted from the application. They are not inventors for the present claims in the application.

Respectfully submitted,

NEW ENGLAND BIOLABS, INC.



Harriet M. Strimpel, D. Phil.
Reg. No. 37,008
Attorney for Applicant
240 County Road
Ipswich, MA 01938
(978) 380-7373

Dated: May 4, 2006

New England Biolabs, Inc.
32 Tozer Road
Beverly, MA 01915

DECLARATION AND POWER OF ATTORNEY Original Application

Atto: Docket No. NEB-166-PUS

As a below named inventor, I hereby declare that:
My residence, post address and citizenship are as stated below next to my name.
I believe that I am the original, first and sole inventor (in only one name is listed at 201 below) or an original, first and joint inventor (if plural names are listed at 201-203 below) of the subject matter which is claimed and which a patent is sought on the invention entitled:

INCORPORATION OF MODIFIED NUCLEOTIDES BY ARCHAEON DNA POLYMERASES AND RELATED METHODS

which is described and claimed in:
[X] the attached specification or [ ] the specification in Application Serial No. _____ filed_____
(for declaration not accompanying application)
And was amended on_____
if applicable

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendments referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a). I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

FOREIGN APPLICATION(S) IF ANY, FILED WITHIN 12 MONTHS PRIOR TO THE FILING DATE OF THIS APPLICATION			
COUNTRY	APPLICATION	FILING (day, month, year)	PRIORITY CLAIMED UNDER 35 U.S.C. 119
			YES NO
			YES NO

ALL FOREIGN APPLICATION(S) IF ANY, FILED MORE THAN 12 MONTHS PRIOR TO THE FILING DATE OF THIS APPLICATION			
COUNTRY	APPLICATION	DATE OF FILING	PRIORITY CLAIMED UNDER 35 U.S.C. 119
PCT	PCT/US00/26900	29-Sep-00	YES

I hereby claim the benefit under Title 35, United States Code §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Serial No.	Filing Date	Status (Patented, Pending, Abandoned)
60/157,204	30-Sep-99	PENDING

POWER OF ATTORNEY:

As a named inventor, I hereby appoint the following attorney with full powers of association, substitution and revocation to prosecute this application and transact all business in the Patent and Trademark Office connected therewith:

Gregory D. Williams
(Registration No. 30901)

SEND CORRESPONDENCE TO:

Gregory D. Williams
General Counsel
New England Biolabs, Inc.
32 Tozer Road
Beverly, MA 01915

DIRECT TELEPHONE CALLS TO:

Gregory D. Williams
General Counsel
New England Biolabs, Inc.
Tele: (978) 927-5054; Ext. 292
Fax: (978) 927-1705

201	Full Name of Inventor	Last Name JACK	First Name WILLIAM	Middle Name E.
	Residence & Citizenship	City WENHAM	State/Foreign Country MASSACHUSETTS	Citizenship US
	Post Office Address	Post Office Address 31 MAYFLOWER DRIVE	City/State/Country WENHAM, MA	Zip Code 01984
202	Full Name of Inventor	Last Name GARDNER	First Name ANDREW	Middle Name F.
	Residence & Citizenship	City BROOKLINE	State/Foreign Country MASSACHUSETTS	Citizenship US
	Post Office Address	PO Address APT.#6 1163 BEACON STREET	City/State/Country BROOKLINE, MA	Zip Code 02446
203	Full Name of Inventor	Last Name BUZBY	First Name PHILIP	Middle Name R.
	Residence & Citizenship	City BROCKTON	State/Foreign Country MASSACHUSETTS	Citizenship US
	Post Office Address	Post Office Address 31 HARLAN DRIVE	City/State/Country BROCKTON, MA	Zip Code 02301
204	Full Name of Inventor	Last Name DiMEO	First Name JAMES	Middle Name J.
	Residence & Citizenship	City NEEDHAM	State/Foreign Country MASSACHUSETTS	Citizenship US
	Post Office Address	Post Office Address 442 CENTRAL AVENUE	City/State/Country NEEDHAM, MA	Zip Code 02494
205	Full Name of Inventor	Last Name	First Name	Middle Name
	Residence & Citizenship	City	State/Foreign Country	Citizenship
	Post Office Address	Post Office Address	City/State/Country	Zip Code

206	Full Name of Inventor	Last Name	First Name	Middle Name
	Residence & Citizenship	City	State/Foreign Country	Citizenship
	Post Office Address	Post Office Address	City/State/Country	Zip Code
207	Full Name of Inventor	Last Name	First Name	Middle Name
	Residence & Citizenship	City	State/Foreign Country	Citizenship
	Post Office Address	Post Office Address	City/State/Country	Zip Code
208	Full Name of Inventor	Last Name	First Name	Middle Name
	Residence & Citizenship	City	State/Foreign Country	Citizenship
	Post Office Address	Post Office Address	City/State/Country	Zip Code
209	Full Name of Inventor	Last Name	First Name	Middle Name
	Residence & Citizenship	City	State/Foreign Country	Citizenship
	Post Office Address	Post Office Address	City/State/Country	Zip Code

I hereby further declare that all statements made herein of my own knowledge are true and th statements made on information and belief are believed to be true and further that these state were made with the knowledge that willful statements and the like so made are punishable by or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that s willful false statements may jeopardize the validity of the application or any patent issued ther

Signature of Inventor 201	Date 18 March 2002
Signature of Inventor 202	Date 18 March 2002
Signature of Inventor 203	Date 3/13/02
Signature of Inventor 204	Date 3/13/02
Signature of Inventor 205	Date
Signature of Inventor 206	Date
Signature of Inventor 207	Date
Signature of Inventor 208	Date
Signature of Inventor 209	Date

Docket No.: NEB-166-PUS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: Jack et al. EXAMINER: Hutson

SERIAL NO.: 10/089,027 ART UNIT: 1652

DATE FILED: March 26, 2002

TITLE: Incorporation of Modified Nucleotides By Archaeon DNA Polymerases And Related Methods

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. §1.131

As a below named inventor, I hereby declare that:

1. My name is Dr. William Jack, Research Director for the DNA Enzymes Division at New England Biolabs Inc. My resume is attached.

2. I have been studying the structure and function of DNA polymerases for over 16 years.

3. I was a member of the group of scientists at New England Biolabs that isolated, characterized, and cloned the first hyperthermophilic archaeal DNA polymerase. Our continuing work with archaeon DNA polymerases identified a surprisingly homogeneous set of enzymes. We claimed this group of DNA polymerases in US Patent 5,500,363. In this patent, the United States Patent and Trademark Office recognized the validity of our claim to a class of archaeon DNA polymerases defined by the DNA encoding the enzyme and its

ability to hybridize under defined conditions to various specified DNA sequences. The group was exemplified by T.litoralis (Vent), GBD (Deep Vent), and 9°N DNA Polymerases.

4. We also found that this group of polymerases had a high degree of amino acid sequence identity. A comparative three-dimensional alignment of members of this group of enzymes showed a high degree of structural conservation, consistent with the observed high degree of primary amino acid sequence identity/similarity. See for example, Vent (Rodriguez, et al., 2000), Tgo (Hopfner, et al., 1999), D. Tok (Zhao, et al., 1999), and KOD (Hashimoto, et al., 2001) DNA Polymerases.

5. The structural equivalence of this group of polymerases is further supported by experiments reported in Example 10 of the above application in which we show that mutation of an analogous residue in Vent and 9°N DNA Polymerases yields enzymes with equivalent acyclonucleotide incorporation efficiencies.

6. We discovered that this group of enzymes is capable of efficiently utilizing acyclonucleotides as substrates. We demonstrated this property using four examples of polymerases within this tightly defined group. Any molecular biologist of ordinary skill in the art would expect from these findings that this property would occur in all members of the enzyme group defined above.

7. Additionally, my colleagues and I have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides, and also for the enhanced incorporation with Vent A488L and 9°N A485L DNA Polymerase mutants. See Gardner, et al. (2004) on page 11841, column 1, paragraph 2 and page 11841, column 2, paragraph

1, respectively.

8. I assert that the combination of the high degree of homogeneity in DNA and amino acid sequences of archaeon DNA polymerases, plus the structural evidence that modification of specific amino acids alters enzyme specificity, would be sufficient to assure a person of ordinary skill in the art that the class of polymerases as defined above will interact with acyclonucleotide substrates as shown in the above application.

9. To further support the above statements, we have conducted additional experiments to confirm that archeon Family B polymerases with an amino acid sequence identity of greater than 30% can utilize acyclonucleotides as a substrate. This data is attached to the present declaration as appendix 1.

9. I further declare under penalty of perjury pursuant to laws of the United States of America that the foregoing is true and correct and that the Declaration was executed by me on:

William E. Jack

Date: 4 May 2006

References:

A. Gardner, C.M. Joyce, W.E. Jack (2004) J. Biol. Chem. 279, 11834-11842.

H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai (2001) J. Mol. Biol. 306, 469-477.

K.P. Hopfner, A. Eichinger, R.A. Engh, F. Laue, W. Ankenbauer, R. Huber, B. Angerer (1999) Proc. Natl. Acad. Sci. USA 96, 3600-3605.

A.C. Rodriguez, H.W. Park, D. Mao, L.S. Beese (2000) J. Mol. Biol. 299, 447-462.

Y. Zhao, D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, J. Kuriyan (1999) Structure 7, 1189-1199.

William Erle Jack
New England Biolabs
240 County Road
Ipswich, MA 01938
(978) 380-7257
(978) 921-1350 (fax)
email: jack@neb.com

RESEARCH INTERESTS

Enzymatic and structural aspects of protein-nucleic acid interactions. Thermostable DNA polymerase kinetics and function.

RESEARCH EXPERIENCE

New England Biolabs (Beverly, MA).

2005-present	Division Head, DNA Enzymes
1987-present	Senior Staff Scientist Research: Kinetic characterization of thermostable DNA polymerases. Creation and characterization of DNA polymerase variants with altered substrate recognition. Over-expression and characterization of restriction and modification enzymes.
2000-present	New England Biolabs Institutional Biosafety Committee Chair

Rockefeller University (NY, NY) Laboratory of Biochemistry and Molecular Biology.

1983-1987	Postdoctoral Fellow in the laboratory of R.G. Roeder. Research: Structural and functional characterization of wild type and mutant forms of *Xenopus* RNA polymerase III transcription factor A. Glucocorticoid hormone-induced transcription enhancement *in vitro*.

Duke University (Durham, NC) Department of Biochemistry.

1977-1983	Graduate Student in the laboratory of P. Modrich. Research: Kinetics and thermodynamics of DNA site location, recognition and cleavage by *Eco*RI endonuclease.

EDUCATION

Doctor of Philosophy (Biochemistry), Duke University, 1983 (Paul Modrich, advisor).
Bachelor of Arts (Chemistry), *Magna Cum Laude*, University of Utah, 1977.

TRAINING

2000	Sixth National Symposium on Biosafety: Prudent Practices for the New Millennium (Conducted by the Centers for Disease Control and Prevention)

PUBLICATIONS

"Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *J. Biol. Chem.* 279, 11834-11842 (2004).

"Acyclic and Dideoxy Terminator Preferences Denote Divergent Sugar Recognition by Archaeon and Taq DNA Polymerases," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.*, **30**, 605-613 (2002).

"The Kinetic Mechanism of *Eco*RI Endonuclease," David J. Wright, **William E. Jack** and Paul Modrich, *J. Biol. Chem.* **274**, 31896-31902 (1999).

"Determinants of Nucleotide Sugar Recognition in an Archaeon DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.* **27**, 2545-2553 (1999).

"Photochemically Initiated Protein Splicing", Sandra N. Cook, **William E. Jack**, Xiaofeng Xiong, Lora E. Danley, Jonathan A. Ellman, Peter G. Shultz, and Christopher J. Noren, *Angew. Chem. Int. Ed. Engl.* **34**, 1629-1630 (1995).

"Three-dimensional Structure of the Adenine-specific DNA Methyltransferase M·TaqI in Complex with the Cofactor S-adenosylmethionine", J. Labahn, J. Granzin, G. Schluckebier, D.P. Robinson, **W.E. Jack**, I. Schildkraut, and W. Saenger, *Proc. Natl. Acad. Sci. USA* **91**, 10957-10961 (1994).

"Expression, Purification, and Crystallization of Restriction Endonuclease *Pvu*II With DNA Containing Its Recognition Site", K. Balendiran, Joseph Bonventre, Roger Knott, **William Jack**, Jack Benner, Ira Schildkraut, and John E. Anderson, *Proteins: Structure, Function, and Genetics* **19**, 77-79 (1994).

"Protein Splicing Elements: Inteins and Exteins-A Definition of Terms and Recommended Nomenclature", Francine B. Perler, Elaine O. Davis, Gary E. Dean, Frederick S. Gimble, **William E. Jack**, Norma Neff, Christopher J. Noren, Jeremy Thorner and Marlene Belfort, *Nucleic Acids Res.* **22**, 1125-1127 (1994).

"Characterization of a DNA Polymerase from the Hyperthermophile Archaea *Thermococcus litoralis*", Huimin Kong, Rebecca B. Kucera and **William E. Jack**, *J. Biol. Chem.*, **268,** 1965-1975 (1993).

"Protein Splicing Removes Intervening Sequences in an Archaea DNA Polymerase", Robert A. Hodges, Francine B. Perler, Christopher J. Noren and **William E. Jack**, *Nucleic Acids Res.* **20**, 6153-6157 (1992).

"Intervening Sequences in an Archaea DNA Polymerase Gene", Francine B. Perler, Donald G. Comb, **William E. Jack**, Laurie S. Moran, Boqin Qiang, Rebecca B. Kucera, Jack Benner, Barton E. Slatko, Donald O. Nwankwo, S. Kay Hempstead, Clotilde K.S. Carlow and Holger Jannasch, *Proc. Natl. Acad Sci. USA* **89**, 5577-5581 (1992).

"Overexpression, Purification and Crystallization of *Bam*HI Endonuclease", **William E. Jack**, Lucia Greenough, Lydia F. Dorner, Shuang-yong Xu, Teresa Strzelecka, Aneel K. Aggarwal and Ira Schildkraut, *Nucl. Acids Res.* **19**, 1825-1829 (1991).

"Nucleotide Sequence of the *FokI* Restriction-Modification System: Separate Strand-Specificity Domains in the Methyltransferase", Mary C. Looney, Laurie S. Moran, **William E. Jack**, George R. Feehery, Jack S. Benner, Barton E. Slatko and Geoffrey G. Wilson, Gene **80**, 193-208 (1989).

"M.FokI methylates adenine in both strands of its asymmetric recognition sequence", David Landry, Mary C. Looney, George R. Feehery, Barton E. Slatko, **William E. Jack**, Ira Schildkraut, and Geoffrey G. Wilson, *Gene* **77**, 1-10.

"Mechanism of specific site location and DNA cleavage by *Eco*RI endonuclease", Brian J. Terry, **William E. Jack** and Paul Modrich, *Gene Amplif Anal.* volume 5. pp. 103-118 (1987)

"Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis", Brian J. Terry, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **260**: 13130-13137 (1985).

Participation of Outside DNA sequences in the EcoRI Endonuclease Reaction Pathway, **William E. Jack**, dissertation (Duke University, Durham, NC) (1983).

"Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences", Brian J. Terry, **William E. Jack**, Robert A. Rubin and Paul Modrich, *J Biol. Chem.* , **258**: 9820-9825 (1983).

"Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence", **William E. Jack**, Brian J. Terry, and Paul Modrich, *Proc Natl. Acad Sci U S A* **79**: 4010-4014 (1982).

"DNA determinants important in sequence recognition by *Eco*RI endonuclease", A-Lien Lu, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **256**: 13200-13206 (1981).

"Structures and Mechanisms of *Eco*RI DNA Restriction and Modification Enzymes", **William E. Jack**, Robert A. Rubin, Andrea Newman and Paul Modrich, *Gene Amplif. Anal.* volume 1, p 165-179 (1981).

US PATENTS

"Use of site-specific nicking endonucleases to create single-stranded regions and applications thereof," **William E. Jack**, Ira Schildkraut, Julie Forney Menin, US Patent 6,660,475, December 9, 2003.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,834,285, November 10, 1998.

"Modified proteins comprising controllable intervening protein sequences or their elements methods of producing same and methods for purification of a target protein comprised by a modified protein," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, Christopher J. Noren, Shaorong S.C. Chong, Eric Adam, Maurice Southworth, US Patent 5,834,247, November 10, 1998.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,500,363, March 19, 1996.

"Modification of protein by use of a controllable intervening protein sequence," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, US Patent 5,496,714, March 5, 1996.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,352,778, October 4, 1994.

"Purified thermostable DNA polymerase obtainable from thermococcus litoralis," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,322,785, June 21, 1994.

doi:10.1006/jmbi.2000.4403 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 306, 469–477

JMB

Crystal Structure of DNA Polymerase from Hyperthermophilic Archaeon *Pyrococcus kodakaraensis* KOD1

Hiroshi Hashimoto[1], Motomu Nishioka[2], Shinsuke Fujiwara[2] Masahiro Takagi[2], Tadayuki Imanaka[3], Tsuyoshi Inoue[1] and Yasushi Kai[1]*

[1]*Department of Materials Chemistry and*

[2]*Department of Biotechnology Graduate School of Engineering, Osaka University 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan*

[3]*Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University Yoshidahonmachi Sakyo-ku Kyoto 606-8501, Japan*

**Corresponding author*

The crystal structure of family B DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 (KOD DNA polymerase) was determined. KOD DNA polymerase exhibits the highest known extension rate, processivity and fidelity. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanisms of those enzymatic features. Structural comparison of DNA polymerases from hyperthermophilic archaea highlighted the conformational difference in Thumb domains. The Thumb domain of KOD DNA polymerase shows an "opened" conformation. The fingers subdomain possessed many basic residues at the side of the polymerase active site. The residues are considered to be accessible to the incoming dNTP by electrostatic interaction. A β-hairpin motif (residues 242-249) extends from the Exonuclease (Exo) domain as seen in the editing complex of the RB69 DNA polymerase from bacteriophage RB69. Many arginine residues are located at the forked-point (the junction of the template-binding and editing clefts) of KOD DNA polymerase, suggesting that the basic environment is suitable for partitioning of the primer and template DNA duplex and for stabilizing the partially melted DNA structure in the high-temperature environments. The stabilization of the melted DNA structure at the forked-point may be correlated with the high PCR performance of KOD DNA polymerase, which is due to low error rate, high elongation rate and processivity.

Keywords: archaea; crystal structure; family B DNA polymerase; "forked-point"; KOD DNA polymerase

Introduction

DNA polymerases are a group of enzymes that use single-stranded DNA as a template for the synthesis of the complementary DNA strand. These enzymes are multifunction, with both synthetic (polymerase) and one or two degradative modes (5′-3′ and/or 3′-5′ exonucleases) and play an essential role in nucleic acid metabolism including the processes of DNA replication, repair and recombination. Many DNA polymerase genes have been cloned and sequenced. Amino acid sequences deduced from their nucleotide sequences can be classified into four major types: *Escherichia coli* DNA polymerase I (family A), *E. coli* DNA polymerase II (family B), *E. coli* DNA polymerase III (family C) and others (family X).[1] Recently, a new family of DNA polymerases has been identified; all members of this family contain five highly conserved motifs, I-V, and several of these polymerases participate in lesion bypass.[2] This family is called the UmuC/DinB family.[3] Family B DNA polymerases include eukaryotic DNA polymerase α, δ, and ε, which are thought to be components of the replisome and to carry out chromosomal DNA replication. Archaeal proteins involved in gene expression, such as those for DNA replication, transcription, and translation, have been found to be similar to those from eucarya. Therefore, the archaeal system of gene expression is a simplified model of the eukaryotic system. In contrast, the

E-mail address of the corresponding author: kai@chem.eng.osaka-u.ac.jp

0022-2836/01/030469–9 $35.00/0

cellular appearance and organization of archaea are more like those of bacteria.

The first crystal structure of a family B DNA polymerase to be obtained was that of bacteriophage RB69 DNA polymerase (RB69 DNA polymerase).[4] The first crystal structure of archaeal DNA polymerase was DNA polymerase from *Thermococcus gorgonarius* (*Tgo* DNA polymerase).[5] The editing complex of RB69 DNA polymerase has been reported,[6] two further crystal structures of archaeal family B DNA polymerases have recently been reported: Tok DNA polymerase from *Desulfurococcus* sp. Tok[7] is 9°N-7 DNA polymerase from *Thermococcus* sp. 9°N-7.[8]

The *Pyrococcus kodakaraensis* KOD1 is a hyperthermophilic archaeon, with an optimum growth temperature of 95 °C.[9] Enzymes produced in KOD1 were reported to be extremely thermostable and to have eukaryotic characteristics.[9] The optimum temperature of KOD DNA polymerase is 75 °C similar to that of DNA polymerase obtained from *Pyrococcus furiosus* (*Pfu* DNA polymerase). KOD DNA polymerase, however, exhibits the higher extension rate (100-130 nucleotides/second) and processivity (>300 bases); five times and ten to 15 times higher than those of *Pfu* DNA polymerase, respectively.[10] Thermostable DNA polymerases are expected to be suitable enzymes for Polymerase Chain Reaction (PCR) KOD DNA polymerase is, therefore, suitable for DNA amplification by such means. Indeed, KOD DNA polymerase is widely used in rapid and accurate PCR systems (TOYOBO Ltd., Japan).

Although structures of three archaeal DNA polymerases have been determined as described above, no structural information relating to elongation rate, processivity or fidelity is provided. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanism of enzymatic features of KOD DNA polymerase, which are the highest extension rate, processivity and fidelity. Here, we report the crystal structure of DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. The three-dimensional structure of this KOD DNA polymerase may provide useful information to clarify the mechanisms for rapid and accurate reaction. In addition, this information may contribute to the improvement of the PCR properties of enzymes already in use such as thermostability, error rate, elongation rate and processivity, or for designing new enzymes for PCR as well as DNA replication by family B DNA polymerases.

Results and Discussion

Overall structure

KOD DNA polymerase has a disk-like shape with dimensions 60 Å × 80 Å × 100 Å and is made up of distinct domains and subdomains: N-terminal (N-ter: 1-130, 327-368, violet), Exonuclease (Exo: 131-326, blue), Polymerase (Pol) domain including the Palm and Fingers subdomains (369-449, 500-587, brown; and 450-499, green, respectively) and the Thumb domain including thumb-1 and thumb-2 subdomains (588-774, red) (Figure 1(a)). The polymerase active site, containing three conserved carboxylates, (Asp404, Asp540 and Asp542) is located in an anti-parallel β-sheet in the Palm subdomain. The exonuclease active site contains two conserved carboxylates (Asp141 and Glu143) and is located in an anti-parallel β-sheet in the Exo domain. The Polymerase and exonuclease active sites on the molecular surface are indicated by *P* and *E*, respectively (see Figure 4). Structural comparisons of archaeal DNA polymerases (KOD, *Tgo* and 9°N-7 DNA polymerases) are shown in Figure 1(b). The structural architectures of the proteins are identical, but the orientation of the domains and subdomains is different. In the case of the KOD DNA polymerase (red), the Thumb domain is shifted to make an "open" conformation and the portion of the Palm domain neighboring the root of the Thumb domain is slightly shifted as a result of the large movement of the Thumb domain in comparison to other archaeal DNA polymerases. Table 1 shows the averaged temperature factors of the domains and subdomains in the crystal structure of KOD DNA polymerase. The value of the Thumb domain was markedly higher than the others. The structures of many residues in the Thumb-2 subdomain are not defined, because the orientation of the subdomain is highly disordered. Therefore, it is thought that the structure of KOD DNA polymerase described here provides information for the DNA-free, most relaxed conformation. The structure of the editing complex of RB69 DNA polymerase revealed that newly synthesized duplex DNA is grasped by the Pol and Thumb domains. Although the orientation of the Thumb domain is potentially highly flexible, the orientation may be fixed when it binds to the primer-template duplex.

Polymerase domain

The Pol domain is made up of the Fingers and Palm subdomains and has an "L-like" shape (Figure 2(a)). The polymerization mechanism has been studied mainly on family A DNA polymerases (Pol-I). A structural basis for a metal-

Table 1. Averaged temperature factors

Domain	Temperature factor (Å²)
N-ter	38.1
Exo	55.7
Pol	
Fingers	49.5
Palm	52.8
Thumb	93.7
Overall	55.9

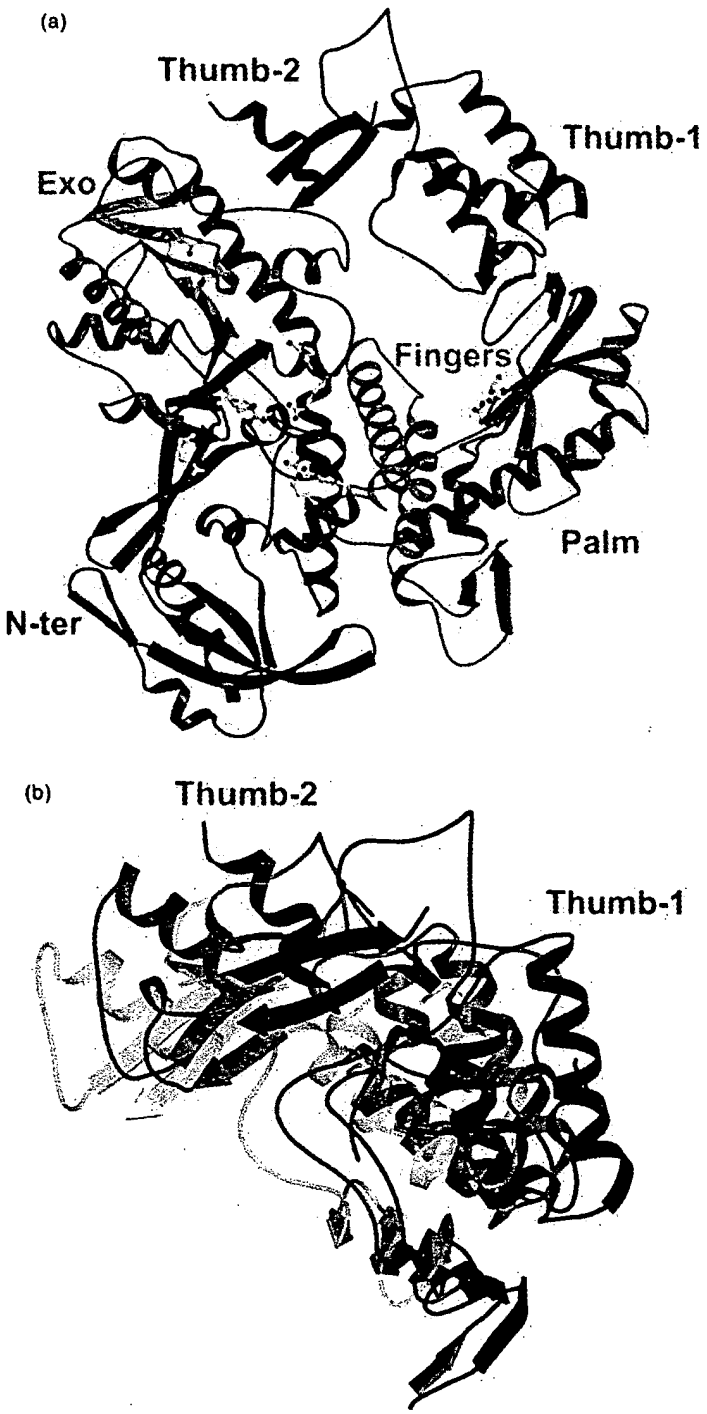


Figure 1. (a) Overall structure of KOD DNA polymerase. The structure is composed of domains and subdomains, which are N-terminal (N-ter, violet), Exonuclease (Exo, blue), Polymerase (Pol) domain including the Palm (brown) and Fingers (green) subdomains and the Thumb domain (red), including the Thumb-1 and Thumb-2 subdomains. Conserved carboxylate residues in Polymerase and Exonuclease active site are shown by ball-and-stick models. (b) Conformational comparison of Thumb domains among three archaeal DNA polymerases. Red, KOD DNA polymerase; blue, *Tgo* DNA polymerase; and green, 9°N-7 DNA polymerase. The comparison shows that the Thumb domain of KOD DNA polymerase displays the most "opened" conformation.

assisted mechanism of phosphoryl transfer was provided by the bacteriophage T7 DNA replication complex.[11] The complex structure shows that two metal ions are bound by strictly conserved carboxylates (Asp475 and Asp654, which correspond to Asp404 and Asp542 in KOD DNA polymerase) extended from the anti-parallel β-sheet of the Palm domain. The phosphate group of incoming ddGTP is held by the metal ions and the four basic residues extending from the Fingers subdomain (His506, Arg518 and Lys522). The crystal structure of two ternary complexes of the large fragment of

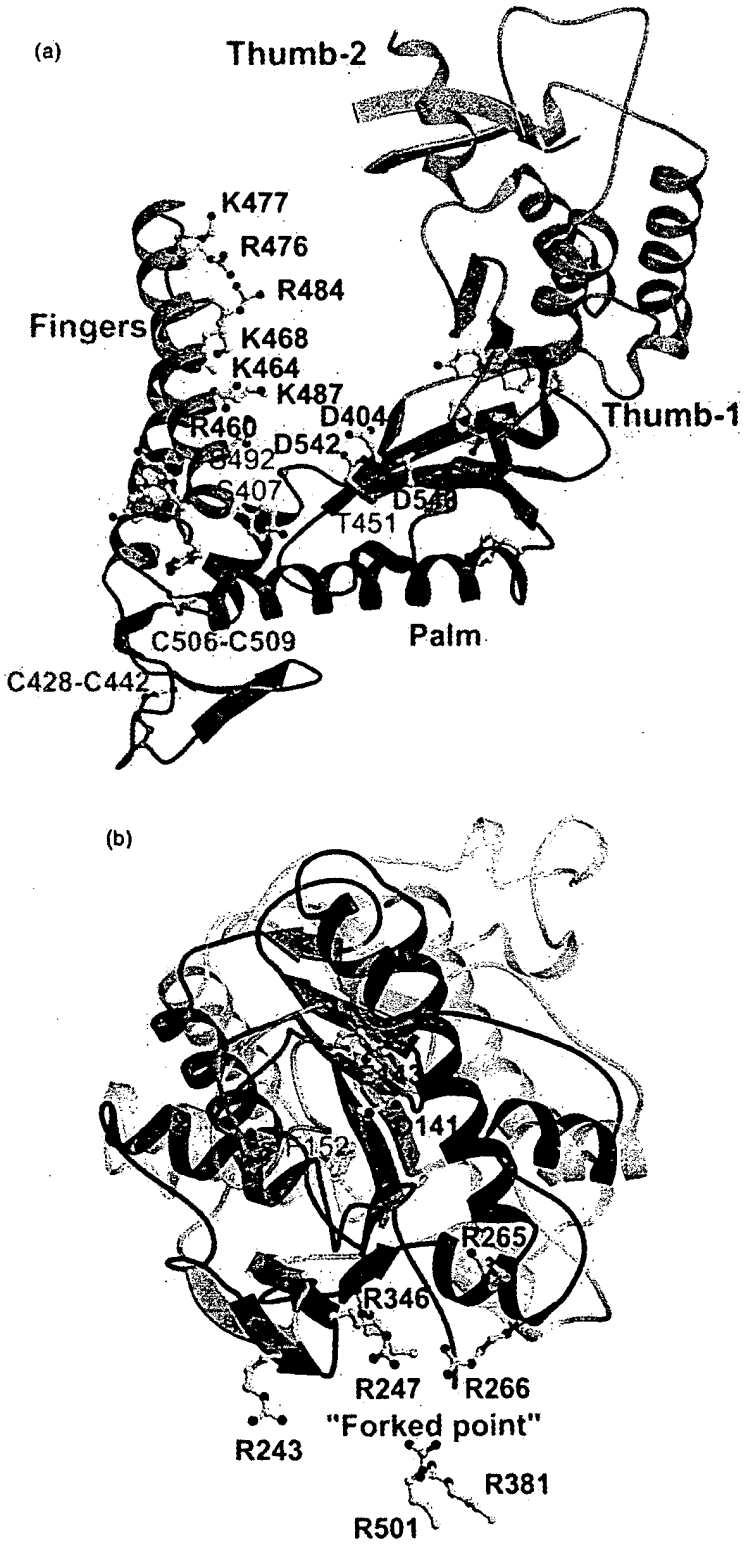


Figure 2. (a) Ribbon representation of the Pol domain. The domain is made up of Fingers and Palm subdomains. Conserved carboxylate residues (D404, D540 and D542) are represented by ball-and-stick models. Basic residues are represented by ball-and-stick models, which stand in a line in the Fingers subdomain, facing the polymerase active site. K464 are replaced by alanine, because of the ambiguity of its electron density. Two disulfide bonds are displayed (C428-C442 and C506-C509). Aromatic residues adjacent to a glycine residue, represented by ball-and-stick models, are localized in the joints of the subdomains. The Thumb domain is represented by the semitransparent model. The C^{α} atoms of the nucleophilic residues, S407 (*Pko* Pol-1), S492 (*Tli* Pol-1) and T541 (*Tli* Pol-2), are represented by violet spheres. (b) Exonuclease domains of KOD DNA polymerase and RB69 DNA polymerase (semitransparent model). Conserved carboxylate (D141 and E143) and arginine residues (R243, R247, R265, R266, R343, R381 and R501) in the forked-point of KOD DNA polymerase are represented by ball-and-stick models. The red strands are β-hairpin motif partitioning template-binding and editing clefts. The loop containing Phe152 is shown in orange. F123 of RB69 DNA polymerase and F152 of KOD DNA polymerase are represented by semitransparent and opaque ball-and-stick models, respectively.

Thermus aquaticus DNA polymerase I (Klentaq1) with a primer-template DNA and ddCTP have been reported.[12] The ternary complexes suggest that basic residues of the Fingers subdomain hold the phosphate group of the incoming dNTP and the domain induces a conformational change to deliver the incoming nucleotide to the active site. In the case of family B DNA polymerases, the Fingers subdomain is composed mainly of two long helices and does not have a joint that appeared in the structures of family A DNA polymerases. Therefore, it seems that in the case of archaeal DNA polymerases, the movement of the Pol domain to deliver dNTP to the active site differs from that of family A DNA polymerases. Kinetic study of RB69 DNA polymerase mutants revealed that four residues (Arg482, Lys486, Lys560 and Asn564) of the Fingers subdomain affected dNTP incorporation.[13] The residues are conserved in family B DNA polymerases, and correspond to Arg460, Lys464, Lys487 and Asn491 in KOD DNA polymerase, respectively. Furthermore, Lys468, Arg476, Lys477 and Arg484 are located at the tip of the Fingers subdomain on the side of the polymerase active site in KOD DNA polymerase (Figure 2(a)). It is expected that the "queue" of basic residues captures the incoming dNTPs, then the dNTP is delivered toward the polymerase active-site center by accompanying the movement of the polymerase domain. Two disulfide bonds exist in the connection site between the Palm and Fingers subdomains (Figure 2(a); Cys428-Cys442 and Cys506-Cys509). The two disulfide bonds are found also in the crystal structures of *Tgo*, Tok and 9°N-7 DNA polymerases. Sequence alignment for archaeal DNA polymerases is shown in Figure 3, suggesting the potential for the formation of disulfide bonds in the same sites. It is thought that the disulfide bonds are required to maintain the structure of the Fingers and Palm subdomains at extremely high temperatures. Sequence comparison suggests that the number of disulfide bonds are correlated with optimum growth temperatures of organisms. DNA polymerases from *Thermococcus litoralis*, *Methanococcus jannaschii* and *Archaeoglobus fulgidus*, with optimum growth temperatures of 85, 85 and 83 °C, respectively, are expected to have one disulfide bond, because Cys506 is replaced by serine in *T. litoralis* and *M. Jannaschii*, and Cys442 is replaced by arginine in *A. fulgidus*. DNA polymerase from *Methanobacterium thermoautotrophicum*, with an optimum growth temperature of 65 °C, is expected to have no disulfide bond, because Cys428, Cys442 and Cys506 are replaced by glutamic acid, arginine and serine, respectively.

Archaeal DNA polymerases have characteristic sequences of aromatic residues adjacent to glycine residues (Figure 3). These are localized at the hinges of the Palm subdomain at the connections to the Fingers and Thumb-1 subdomains (Figure 2(a)). These aromatic residues may provide a flexible aromatic environment because of the adjoining glycine residues. This may contribute to the conformational changes of Pol domain in polymerization.

The 3′-5′ exonuclease domain

DNA is synthesized by competition between the rate of polymerase and exonuclease activities at the newly synthesized 3′ terminus from the primer. Misincorporation of a nucleotide destabilizes the structure of duplex DNA at the 3′ terminus of the primer. This decreases the rate of nucleophilic attack on the α-phosphate group of the incoming dNTP by the primer 3′-OH and allows excision of the incorrect nucleotide by the proofreading exonuclease. The excision requires the movement of the 3′ terminus to the exonuclease active site accompanied by rewinding of the duplex DNA, because the exonuclease active site is set apart from the polymerase active site. In KOD DNA polymerase, the exonuclease active site is set apart from the polymerase active site by approximately 40 Å. The editing complex of RB69 DNA polymerase shows structural similarity to the editing mode of family B DNA polymerase.[6] The DNA polymerase binds the mismatched primer-template DNA, which is partially denatured; the 3′ end of the primer strand is bound at the exonuclease site. Residues 251-262 of RB69 DNA polymerase, that form an extended β-hairpin structure that juts directly out from the protein surface and projects into the DNA, stabilize the partially denatured or melted structure. Arg260 extending from the β-hairpin motif plays an important role. Arg260 and Phe123 appear to block the template strand by making interactions with the penultimate base at the 3′ end of the primer-template. Arg260 and Phe123 in RB69 DNA polymerase correspond to Arg247 and Phe152, in KOD DNA polymerase respectively. Figure 2(b) shows the structural comparison of Exo domains of KOD and RB69 DNA polymerases. Molecular surface and electrostatic potentials are shown in Figure 4. The β-hairpin motif in KOD DNA polymerase corresponds to residues 242-249 and Arg247, extending to the forked-point, which is the junction of the template-binding and editing clefts (T-cleft and E-cleft, respectively) (Figure 4). It seems that Arg247 can separate template strand from primer strand and stabilize the melted structures of the strands in a manner similar to that of the RB69 DNA polymerase. As Phe152 is set apart from the active site, it is apparently unable to make an aromatic interaction with the base of the primer. Based on the above idea, the movement of the loop including Phe152 (Figure 2(b)) is required to interact with the primer bound at the E-cleft. Furthermore, Arg243 extends from the β-hairpin structure to the T-cleft. Arg243 interacts with the template strand to fix it at the T-cleft. In addition to Arg243 and Arg247, five arginine residues gather at the forked-point in KOD DNA polymerase (Arg265, Arg266, Arg346, Arg381 and Arg501) and provide a basic environment (Figures 2(b) and 4). It seems that they can interact with the phosphate

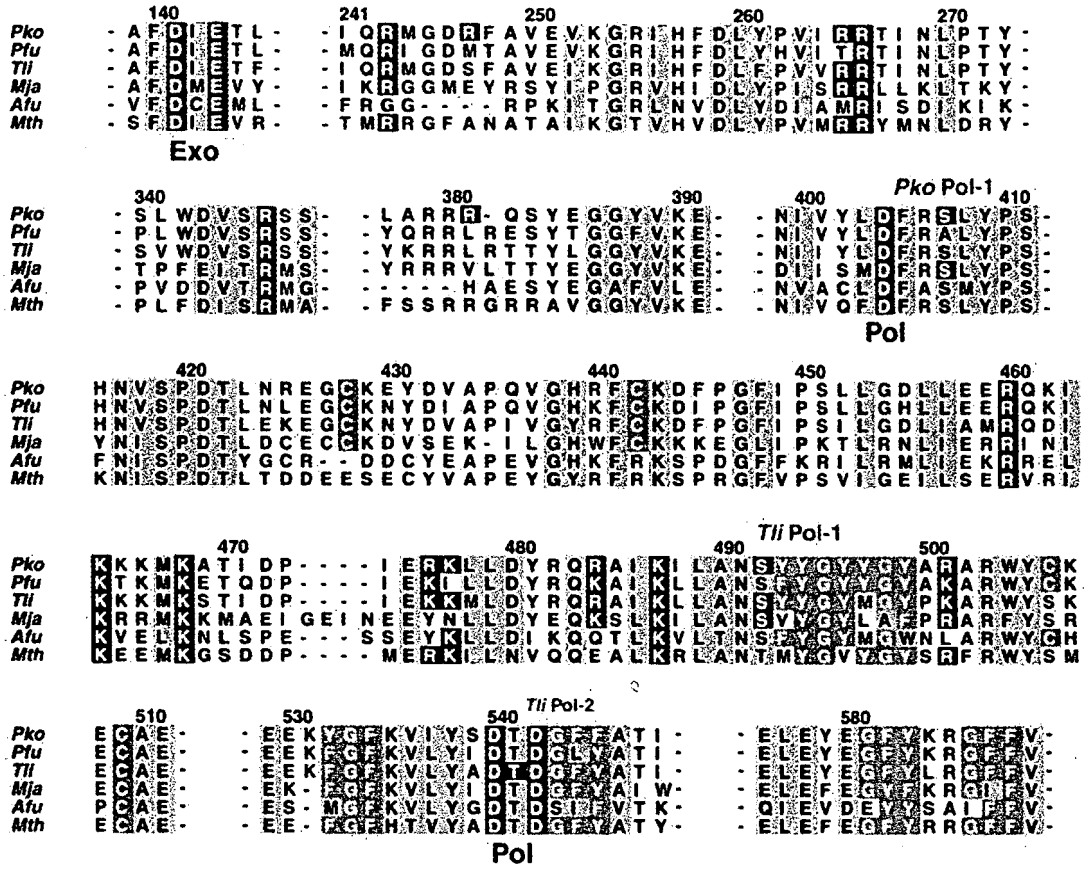


Figure 3. Sequence alignment of archaeal DNA polymerases. The abbreviations used as follows: *Pko, Pyrococcus kodakaraensis; Pfu, Pyrococcus furiosus; Tli, Thermococcus litoralis; Mja, Methanococcus jannaschii; Afu, Archaeoglobus fulgidus;* and *Mth, Methanobacterium thermoautotrophicum.* Homologous residues are masked in gray. Remarkable residues are highlighted in reverse type. Conserved carboxylate residues in the Exonuclease and Polymerase active sites are shown in red. Basic residues gathering in the forked-point and Fingers subdomain are shown in blue. R243, R247, R255, R266, R346, R381 and R501 are located in the forked-point. R460, K464, K468, R476, K477, R484 and K487 are located in the Fingers subdomain and face into the polymerase active site. Cysteine residues forming (or possibly forming) disulfide bonds are shown in green. Nucleophilic residues in self-splicing reaction are shown in violet. Inteins intervene before the nucleophilic residues. Aromatic residues adjacent to glycines are shown in orange.

groups of the DNA strand and stabilize the melted structure of DNA strands at the forked-point. Several arginine residues at the forked-point are conserved in known family B DNA polymerases from hyperthermophilic archaea.

In DNA synthesis, the structure of DNA is variable at the stage of switching between the elongation and editing modes. Hypethermophiles must have mechanisms to protect their genomic DNA against thermal denaturation. The genomic DNA of hyperthermophilic archaea have nucleosome-like structures brought about by interaction with histone-like proteins.[14] Nevertheless, at the replication fork, the DNA strands are exposed. Therefore, DNA polymerases of hyperthermophilic archaea are required to stabilize the exposed or melted DNA structure in the high temperature environment. The stabilization by DNA polymerse may correlate with the enzymatic characteristics of DNA polymerase such as half-life period of activity, error rate, elongation rate and processivity. As discussed above, it is considered that the arginine residues around the "forked-point" have a remarkable effect on the stability of DNA structure. In the forked-point of *Pfu* DNA polymerase, Arg247, Arg265 and Arg501 are replaced by methionine, threonine and lysine, respectively. Therefore, the replacements may affect the difference of the enzymatic characteristics between KOD and *Pfu* DNA polymerases. Additional experiments such as site-directed mutagenesis, a together with enzymatic studies of DNA polymerases are necessary to clarify the role of the residues at the forked-point.

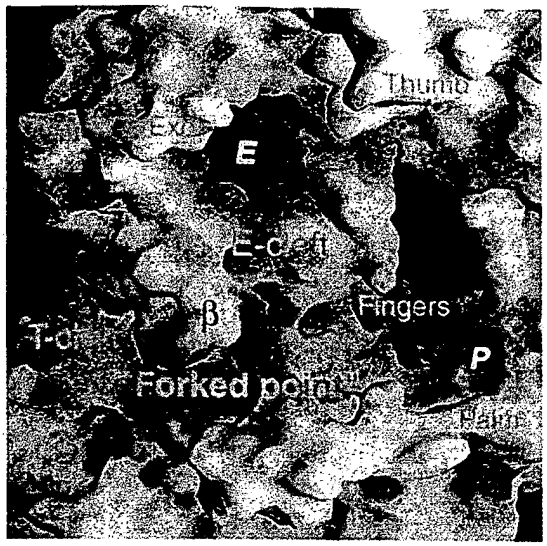


Figure 4. Molecular surface with electrostatic potential map around the forked-point. The red and blue surfaces are acidic and basic regions, respectively. Domains and subdomains are labeled with orange letters. Polymerase and Exonuclease active sites are labeled with P and E, respectively. The β-hairpin is labeled with β.

Extein connection site

The KOD DNA polymerase gene encodes a 1671 amino acid residues precursor protein. The precursor protein is processed precisely into three parts by protein splicing. The self-splicing reaction yields the mature KOD polymerase (774 residues and two intervening protein domains (termed inteins), PI-*Pko*I (360 residues) and PI-*Pko*II (537 residues) as a result of the ligation of the external N and C-terminal domains (termed extein).[10,15] All known precursor proteins contain conserved amino acids at self-splicing sites: serine, threonine or cysteine (nucleophiles) at the intein N terminus, and His-Asn pair at the intein C terminus followed by serine, threonine, cysteine (nucleophiles) at the C-extein N-terminus.[16] The traces of the protein splicing reaction in KOD DNA polymerase are Ser407 and Ser492, which were located in at the N terminus of the C-extein. In the crystal structure of KOD DNA polymerase, the nucleophilic residues are found in the Pol domain (Figure 2(a)).

Self-splicing sites in archaeal family B DNA polymerases (α family) are classified into three types: *Pko* Pol-1, *Tli* Pol-1 and *Tli* Pol-2 (The Intein Database, http://www.neb.com/neb/inteins.html). The nucleophilic residues, serine or threonine, in the three sites are mapped in Figure 2(a). In the case of KOD DNA polymerase, PI-*Pko*I intervenes in the *Pko* Pol-1 site and PI-*Pko*II intervenes in the *Tli* Pol-1 site. The structure shows that they are localized around the polymerase active site in the Palm domain. Although they are exposed to solvent, they are surrounded by the Fingers subdomain and the Thumb domain. The two inteins cannot exist in the space because of steric hindrance. Therefore, it is necessary that the folding of inteins and the subsequent self-excisions are carried out before the extein is folded.

Materials and Methods

Crystallization

KOD DNA polymerase was overexpressed in *E. coli* BL21(DE3) and purified by the previously reported method.[10] The crystals of KOD DNA polymerase were grown by the previously reported method.[17]. KOD DNA polymerase was concentrated up to about an $A_{280\ nm}$ of 25. Crystals of KOD DNA polymerase suitable for diffraction experiments were obtained at 293 K with hanging drops of 2 μl of protein solution and 2 μl of reservoir solution containing 100 mM sodium citrate buffer (pH 5.5) and 25 ~ 30% (v/v) 2-methyl-2,4-pentanediol (MPD), equilibrated against the reservoir solution.

Data collection

X-ray diffraction measurements were performed at the beamline 18B of the Photon Factory at the High Energy Accelerator Research Organization, Tsukuba Science City, Japan. Each crystal of KOD DNA polymerase was picked up directly with a nylon fiber loop from a drop of mother liquid; the crystal was then rapidly transferred to the N_2 gas stream. The incident beam with wavelength of 1.00 Å was collimated to 0.2 mm in diameter. Intensity data were collected on 200 mm × 400 mm imaging plates (Fuji Film Company Ltd.) using the Weissenberg camera for macromolecules with a radius of 430 mm[18,19] and the oscillation method with 3° rotation per frame. The crystals diffracted at least to 2.8 Å resolution at 100 K. X-ray diffraction data were processed and scaled with programs DENZO and SCALEPACK.[20] The diffraction data were scaled with zero σ cutoff. Unit-cell parameters were determined as $a = 111.9$ Å, $b = 112.4$ Å and $c = 73.9$ Å with the space group of $P2_12_12_1$. The unit-cell parameters gave Matthew's coefficient of 2.60 $Å^3Da^{-1}$ and a solvent content of 52.2% (v/v).[21] The final completeness of the data consisted of 119,205 measurements of 20,298 unique observed reflections with an overall R_{merge} of 8.4% and 34.5% in the outermost resolution shell (2.90-2.80 Å). This represents 88.1% of theoretically observable reflections at 2.8 Å resolution. The outermost resolution shell of data is 83.7% complete.

Structure determination

The crystal structure of KOD DNA polymerase was solved by molecular replacement with the AMoRe program.[22] The structure of *Tgo* DNA polymerase (PDB code 1TGO) reduced to polyalanine was used as the search model. Data in the resolution range of 20.0-3.5 Å were used in both the rotation and translation functions. Results are discussed in terms of the AMoRe correlation coefficient (CC). Using a Patterson cut-off radius of 36 Å, a list of 20 rotation function peaks was obtained, with the top peak having an AMoRe CC value of 13.8. The top solution by translation function is CC of 43.3 with an *R*-factor of 54.1%. At this stage, the electron density of the Thumb domain is very ambiguous. Therefore, structural refinement of the initial stage was carried out with

Figure 5. The final $2F_o - F_c$ map around the Fingers and Palm subdomains. The map is contoured at 1 σ.

a model omitting the Thumb domain. The model was manually modified using the program O[23] and subjected to further rounds of refinement using data in the resolution range 40.0-3.0 Å with the program CNS.[24] The final *R*-factor is 23.1% and R_{free} is 31.3%, with r.m.s. deviations for bond lengths and bond angles being 0.007 Å and 1.1°, respectively. The 50 residues at the tip of one Thumb domain are not included in the final model due to poorly defined electron density. Figure 5 shows the final $2F_o - F_c$ map superimposed on the refined final coordinates of KOD DNA polymerase.

Protein Data Bank accession code

Refined coordinates and structure factor have been deposited in the RCSB Protein Data Bank under the accession code 1GCX.

Figure preparation

Figures 1 and 2 were prepared using programs MOLSCRIPT[25] and Raster3D.[26,27] Figure 4 was prepared by GRASP.[28] Figure 5 was prepared using the program O.[23]

Acknowledgments

We thank Professor N. Sakabe, Dr N. Watanabe, Dr M. Suzuki and Dr N. Igarashi, for support in data collection at KEK-PF, Japan. This study was supported by TARA Sakabe Project of University of Tsukuba. The author is grateful for a JSPS Fellowship for Japanese Junior Scientists.

References

1. Braithwaite, D. K. & Ito, J. (1993). Compilation alignment, and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.* **21**, 787-802.
2. Johnson, R. E., Washington, W. T., Prakash, S. & Prakash, L. (1999). Bringing the gap: a family of novel DNA polymerases that replicate faulty DNA. *Proc. Natl Acad. Sci. USA*, **96**, 12224-12226.
3. Friedberg, E. C., Feaver, W. J. & Gerlach, V. L. (2000). The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance. *Proc. Natl Acad. Sci. USA*, **97**, 5681-5683.
4. Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steiz, T. A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. *Cell*, **89**, 1087-1099.
5. Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl Acad. Sci. USA*, **96**, 3600-3605.
6. Shamoo, Y. & Steiz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell*, **99**, 155-166.
7. Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R. & Kuriyan, J. (1999). Crystal structure of an archaebacterial DNA polymerase. *Structure*, **7**, 1189-1199.
8. Rodriguez, A. C., Park, H.-W., Mao, C. & Beese, L. S. (2000). Crystal structure of a pol α family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. 9°N-7. *J. Mol. Biol.* **299**, 471-487.
9. Fujiwara, S., Okuyama, S. & Imanaka, T. (1996). The world of archaea: genome analysis, evolution and thermostable enzymes. *Gene*, **179**, 165-170.
10. Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. & Imanaka, T. (1997). Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* **63**, 4504-4510.
11. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, **391**, 251-258.
12. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.
13. Yang, G., Lin, T.-C., Karam, J. & Konigsberg, W. H. (1999). Steady-state kinetic characterization of RB69 DNA polymerase mutants that affect dNTP incorporation. *Biochemistry*, **38**, 8049-8101.
14. Grayling, R. A., Sandman, K. & Reeve, J. N. (1996). DNA stability and DNA binding proteins. *Advan. Protein. Chem.* **48**, 437-467.
15. Nishioka, M., Fujiwara, S., Takagi, M. & Imanaka, T. (1998). Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of *Pyrococcus kodakaraensis* strain KOD1. *Nucl. Acids Res.* **26**, 4409-4412.
16. Perler, F. B., Olsen, G. J. & Adam, E. (1997). Compilation and analysis of intein sequences. *Nucl. Acids Res.* **25**, 1087-1093.
17. Hashimoto, H., Matsumoto, T., Nishioka, M., Yuasa, T., Takeuchi, S., Inoue, T., Fujiwara, S., Takagi, M.,

Imanaka, T. & Kai, Y. (1999). Crystallographic studies on a family B DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1. *J. Biochem.* **125**, 983-986.
18. Sakabe, N., Ikemizu, S., Sakabe, K., Higashi, T., Nakagawa, A., Watanabe, N., Adachi, S. & Sasaki, K. (1995). Weissenberg camera for macromolecules with imaging plate data collection system at the photon factory, present status and future plan. *Rev. Sci. Instrum.* **66**, 1276-1281.
19. Watanabe, N., Nakagawa, A., Adachi, S. & Sakabe, N. (1995). Macromolecular crystallography station BL-18B at the photon factory. *Rev. Sci. Instrum.* **66**, 1824-1826.
20. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
21. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
22. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157-163.
23. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1994). Improved methods for building protein models in electron density maps and the location of errors in the model. *Acta Crystallog. sect. A*, **47**, 110-119.
24. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
25. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
26. Merritt, E. A. & Murphy, M. E. (1994). Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869-873.
27. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.
28. Nicholls, A., Sharp, K. A. & Honig, B. (1991). GRASP. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

Edited by R. Huber

(Received 21 August 2000; received in revised form 8 December 2000; accepted 8 December 2000)

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 3600–3605, March 1999
Biochemistry

Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*

(x-ray structure/disulfide bonds/replication/Archaea/exonuclease)

KARL-PETER HOPFNER*†‡, ANDREAS EICHINGER*, RICHARD A. ENGH*§, FRANK LAUE§, WALTRAUD ANKENBAUER§, ROBERT HUBER*, AND BERNHARD ANGERER§

*Abteilung Strukturforschung, Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany; and §Roche Diagnostics, D-82372 Penzberg, Germany

Contributed by Robert Huber, January 22, 1999

ABSTRACT **Most known archaeal DNA polymerases belong to the type B family, which also includes the DNA replication polymerases of eukaryotes, but maintain high fidelity at extreme conditions. We describe here the 2.5 Å resolution crystal structure of a DNA polymerase from the Archaea *Thermococcus gorgonarius* and identify structural features of the fold and the active site that are likely responsible for its thermostable function. Comparison with the mesophilic B type DNA polymerase gp43 of the bacteriophage RB69 highlights thermophilic adaptations, which include the presence of two disulfide bonds and an enhanced electrostatic complementarity at the DNA–protein interface. In contrast to gp43, several loops in the exonuclease and thumb domains are more closely packed; this apparently blocks primer binding to the exonuclease active site. A physiological role of this "closed" conformation is unknown but may represent a polymerase mode, in contrast to an editing mode with an open exonuclease site. This archaeal B DNA polymerase structure provides a starting point for structure-based design of polymerases or ligands with applications in biotechnology and the development of antiviral or anticancer agents.**

Propagation of cells requires faithful DNA replication. This is performed *in vivo* by DNA polymerases (pols), which attach the appropriate dNTP to the nascent DNA primer strand to match its paired template. Different families of pols are involved in different DNA polymerization processes including not only DNA replication (1, 2) but also repair and recombination (3, 4), a heterogeneity also reflected by varying polypeptide structures and/or subunit compositions (3, 5). Some pols complement polymerase activity with 3′ → 5′ exonuclease activity (editing activity) and/or 5′ → 3′ "structure-specific endonuclease" activity, often located in separate structural domains on the same polypeptide chain (4–8).

Crystal structures are available for most known polymerase families, including the A family DNA polymerases (9–14), pol β (15–17), HIV reverse transcriptase (18–20), and recently, the B family pol gp43 from bacteriophage RB69 (21). All share a functional polymerase structure, which resembles a right hand built by the palm, fingers and thumb domains (see ref. 7 for review). Although the fingers and thumb domains are highly diverse among the different families, the palm domains, which contain the conserved catalytic aspartate residues, show a similar topology among all families except pol β. The polymerase nucleotidyl transfer was studied in detail for the A family polymerases, HIV reverse transcriptase, and pol β, and was shown to involve two metal ions (summarized in ref. 7).

Considerably less is known for the family of type B pols, which are replicative enzymes in eukaryotes and most likely also Archaea (22, 23). The structure of gp43 from bacteriophage RB69 (21) provided an excellent first insight into this family. In addition to the three polymerase domains, gp43 contains an 3′ → 5′ exonuclease domain and an N-terminal domain. The exonuclease and palm domains share the topology and active site of A family enzymes, implying similar metal-assisted mechanisms for polymerase and exonuclease activities (21). The thumb and finger domains are apparently unrelated to the other polymerase families. The function of the N-terminal domain remains unknown, but may help assemble the multicomponent replication apparatus (21).

Much is known about the replication of phages (24–26), viruses (1, 27), Prokaryota, (28) and Eukaryota (1, 3, 29, 30), which in general involves pols but also primases, helicases, RNaseH, sliding clamps, and other factors (31). Considerably less is known for archaeal replication, where mostly B type polymerases, similar to eukaryotic replication enzymes pol α and δ, have been identified (6, 22, 23, 32–34). This relative ignorance is surprising, because such crucial biotechnological applications as cloning and PCR require the thermostability and fidelity typical of archaeal polymerases (6). Thus, in addition to satisfying basic research interests, structural information could assist, for example, the engineering of variant enzymes with tailored nucleotide incorporation rates or the design of antiviral and anticancer polymerase inhibitors. For these reasons, we have determined the structure of a DNA polymerase from *Thermococcus gorgonarius* (*Tgo*), an extremely thermophilic sulfur-metabolizing archaeon isolated from a geothermal vent in New Zealand (35). This enzyme possesses pol and a 3′ → 5′ exonuclease activity, which together ensure thermostable replication with high fidelity (error rate: 3.3–2.2 × 10^{-6}; see ref. 36). The 2.5 Å structure shows a topological similarity to gp43 and gives insight in the structural biology of archaeal DNA polymerases, including the identification of several mechanisms for thermophilic adaptation.

MATERIALS AND METHODS

Materials. All materials were of the highest grade commercially available.

Bacterial Strains. *Escherichia coli* LE392 containing pUBS520 was used as described (36). *E. coli* B834 (DE3) (hsd metB) was a generous gift of Nediljko Budisa (Max-Planck-Institut).

Expression Vectors. PBTac2 was obtained from Roche Molecular Biochemicals.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: *Tgo*, *Thermococcus gorgonarius*; pol, DNA polymerase.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, Biology Department, Brookhaven National Laboratory, Upton, NY, 11973 (PDB ID code 1TGO).

†Present address: The Scripps Research Institute, La Jolla, CA 92037.
‡To whom reprint requests should be addressed. email: hopfner@scripps.edu.

Table 1. Data collection and isomorphous replacement

Derivative	Resolution limit, Å	Total observations	Unique observations	Completeness, %	R_{sym}, %	R_{iso}, %	Phasing power
Native	3.0	136,953	21,529	91.9	7.0		
U	3.2	25,981	16,119	89.5	11.9	18.1	0.31
PT1	3.7	65,464	11,692	98.2	12.0	22.6	1.54
PT2	4.0	53,870	6,293	66.9	13.3	13.8	1.77
PT3	3.4	85,226	14,155	92.8	10.5	17.1	1.16
PT4	2.6	387,925	27,487	90.6	5.7	35.5	0.67
PT5	2.7	358,695	24,543	88.7	6.1	36.2	0.83
PTU	3.5	79,629	11,437	82.4	9.4	24.7	1.54
PB	3.5	78,190	11,336	92.9	11.5	12.8	0.41
PBPT	3.4	94,557	14,129	91.6	8.0	19.6	1.04
OS	2.8	600,108	24,026	91.2	5.3	38.4	1.82

Overall figure of merit (15.0–3.0 Å): 0.73

Heavy-atom derivatives were prepared by soaking the crystals in low-salt buffer containing the heavy atom as follows: U, 0.5 mM uranyl acatate for 2 h; PT1, 5 mM K_2PtCl_4 for 1 d, PT2, 5 mM K_2PtCl_6 for 2 d; PT3, saturated *cis*-dichlorodipyridine-Pt(II) for 2 d; PT4, 5 mM K_2PtCl_6 for 7 d; PT5, saturated *cis*-dichlorodipyridine-Pt(II) for 7 d; PTU, PT1 + U; PB, saturated dinitrophenyl-$Pb(NO_3)_2$ for 7 d; PBPT, PB + PT5; OS, saturated K_2OsO_4 for 7 d. NAT1, PT1, PT2, PT3, PTU, PB and PBPT were collected with a Mar imaging plate and U was collected with a Bruker AXS area detector on a Rigaku rotating anode source. All other data sets were collected with a Mar charge-coupled device at beamline BW6 at DESY, Hamburg.

Cloning, Expression, and Protein Purification. The gene for the 89.8-kDa DNA-dependent pol (*Tgo* pol) was cloned from *Tgo* (Deutsche Sammlung von Mikroorganismen no. 8976) and expressed in *E. coli* LE392pUBS520 pBtac2*Tgo* (Deutsche Sammlung von Mikroorganismen no. 11328) as described (36). *Tgo* pol was purified essentially as described (36) with the substitution of the TSK butyl Toyopearl column by Blue-Trisacryl M (Serva) and with an additional concentration step on Poros 50 HQ anion exchange medium (Roche Molecular Biochemicals). Active fractions were combined, concentrated to 30 mg/ml, and transferred to 20 mM sodium phosphate, pH 8.2/10 mM 2-mercaptoethanol/500 mM NaCl.

The gene for a selenomethionine-containing variant of *Tgo* pol (Se-*Tgo* pol) was expressed in *E. coli* B842 (DE3) (hsd metB) using a published protocol (37). Se-*Tgo* pol was purified by using the wild-type protocol.

Crystallization. Crystals of purified *Tgo* pol (or Se-*Tgo* pol) were grown by using sitting-drop vapor-diffusion technique at 4°C with high-salt conditions (2:2 μl protein:reservoir—100 mM Tris, pH 8/2.0M ammonium sulfate) and diffracted to 3.0 Å (in-house) and to 2.5 Å [beamline BW6 at Deutsches Elektronen Synchrotron (DESY), Hamburg]. Low-salt conditions (100 mM Tris, pH 7.0/200 mM ammonium sulfate/30% PEG 400) yielded only poorly diffracting crystals but allowed soaks (including heavy atoms) with some cell constant modulation (a, b, c = 63.6, 105.0, 160.5) but minimal loss of resolution.

Data Collection and Processing. Data were collected with a MAR imaging plate or a Bruker AXS X1000 mounted on a Rigaku rotating anode source, or with a MAR imaging plate or a MAR CCD (charge-coupled device) at beamline BW6 at DESY, Hamburg. The data were processed with SAINT (Bruker AXS), MOSFLM (Mar CCD; ref 38), or DENZO (MAR imaging plate; ref 39), scaled with SCALA (40) or SCALEPACK (39), and reduced with TRUNCATE (40).

Structure Determination. The structure was solved by multiple isomorphous replacement and anomalous scattering (MIRAS) by using data from crystals transferred to low-salt conditions (Table 1). Crystallographic calculations were done with programs from the CCP4 suite (40). Heavy atom positions of major sites were located in difference Patterson maps and were refined with MLPHARE (40) to calculate protein phase angles to 3.5 Å resolution. A partial polyalanine model was built into interpretable portions of secondary structural elements of the MIRAS map by using MAIN (41). The quality of the electron density was improved by phase combination of the partial model with the experimental phases by using SIGMAA (40), and several cycles of solvent flattening to 3.0 Å by using SOLOMON (40). At this stage, no interpretable density was found for a significant portion of the molecule, comprising residues 147–154, 283–306, 653–728, and 752–773.

Model Building and Refinement. The partial model (R factor 35%) was used to phase the 2.5 Å resolution data of the Se-*Tgo* pol (high-salt conditions). The model was oriented with AMORE (40). The correlation coefficient of 22.0% and the R factor of 50.3% showed divergence of the high- and low-salt structures. After bulk solvent correction, anisotropic B factor correction and rigid-body minimization (treating five domains independently), the partial model was iteratively refined and extended with simulated annealing, Powell minimization, restrained individual B factor refinement with CNS (42), and manual model building with MAIN (41) by using data from 25.0–2.5 Å resolution (Table 2, Fig. 1).

RESULTS AND DISCUSSION

Structure of *Tgo* pol. *Tgo* pol is a ring shaped molecule with dimensions 50 Å × 80 Å × 100 Å. The single polypeptide chain of 773 aa is folded into five distinct structural domains (Fig. 2): the N-terminal domain (residues 1–130), the 3′ → 5′ exonuclease domain (131–326), the palm (369–449 and 500–585), fingers (450–499), and thumb (586–773) domains of the polymerase unit, and a helical interdomain insertion (327–368)

Table 2. Crystallographic refinement, high-salt form

Space group	$P2_12_12_1$
Cell dimensions, Å	a = 58.1, b = 105.2, c = 154.2
Observations, 25–2.5 Å	
Total	482,448
Unique	30,451
Completeness, %	
Total	91.1
Last shell	86.0
R_{sym}, %	
Total	7.1
Last shell	26.2
R factor (R_{free}), %	20.9 (27.1)
rms deviation in bond lengths, Å	0.008
rms deviation in bond angles, °	1.5
No. of nonhydrogen atoms	
Protein	6,328
Water	339

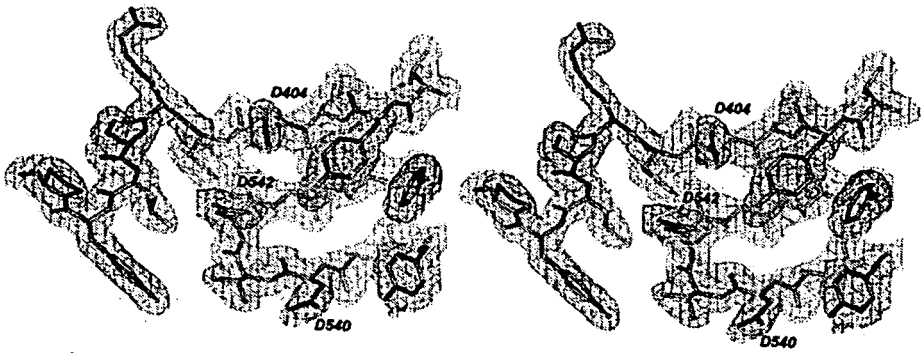


FIG. 1. Stereorepresentation of the electron-density map. The 2 ($F_o - F_c$) electron density contoured at 1.0 σ at the polymerase active site is well defined for the refined model (stick representation).

between the exonuclease and palm domains. The polymerase unit forms the DNA-binding crevice, reminiscent of a right hand, which is the identifying characteristic of pols. Gp43 from bacteriophage RB69 also shows this overall domain topology (21).

Three clefts extend radially from the polymerase active site at the center of the ring: two of them in opposite directions, forming a large cleft across the molecule, and one perpendicular to these. Based on active-site homology to pol A family enzymes, the two opposite clefts probably bind duplex DNA (cleft D, according to ref. 21) and single-strand template DNA (cleft T), respectively. The perpendicular (editing) cleft links the polymerase active site and the exonuclease active site and binds the primer strand in editing mode (21).

The exonuclease domain is structurally equivalent to the 3′ → 5′ exonuclease domain of pol A family (43). Like gp43, however, it is bound at the opposite side of the polymerase unit by noncovalent contacts to the thumb domain at the editing cleft, on one side, and by covalent and noncovalent contacts to the N-terminal and palm domains and the 42 residue interdomain helix, on the other side. This latter segment is located at the base of cleft T, which is additionally bounded by the exonuclease, N-terminal, and palm domains.

The topology of the palm domain is conserved among polymerase families (5), with two long helices (Q and R)

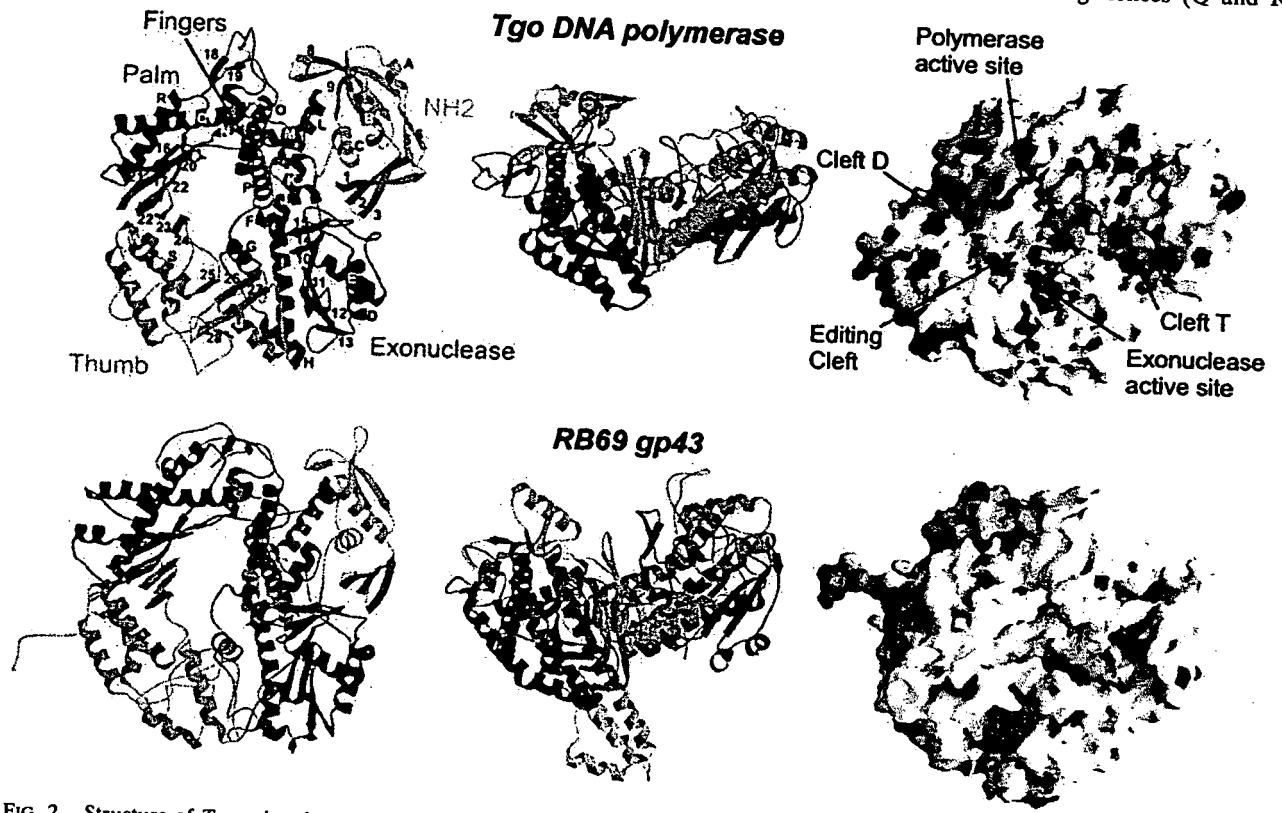


FIG. 2. Structure of *Tgo* pol and comparison with gp43 form bacteriophage RB69. (*Left* and *middle*) Ribbon representation of *Tgo* DNA polymerase (*Upper*) in two orthogonal views with labeled secondary structure elements. The molecule is composed of five domains: N-terminal domain (yellow), 3′ → 5′ exonuclease (red), palm [light and dark magenta represent the N-terminal and C-terminal segment, respectively (see text)], fingers (blue), and thumb (green), which are arranged to form a ring. An interdomain helical segment between the exonuclease domain and the palm is orange. The conserved carboxylates in the active site and the two disulfide bridges are shown as magenta and yellow ball-and-sticks, respectively. *Tgo* pol has the same overall architecture and domain topology than gp43 of RB69 (*Lower*). The 50-residue insertion in the fingers of gp43 is gray. (*Right*) Comparison of molecular surfaces of *Tgo* pol (*Upper*) and RB69 gp43 (*Lower*). Red and blue denote negative and positive electrostatic surface potentials, respectively. In contrast to gp43, *Tgo* pol has a strongly enhanced positive potential at the putative DNA-binding clefts.

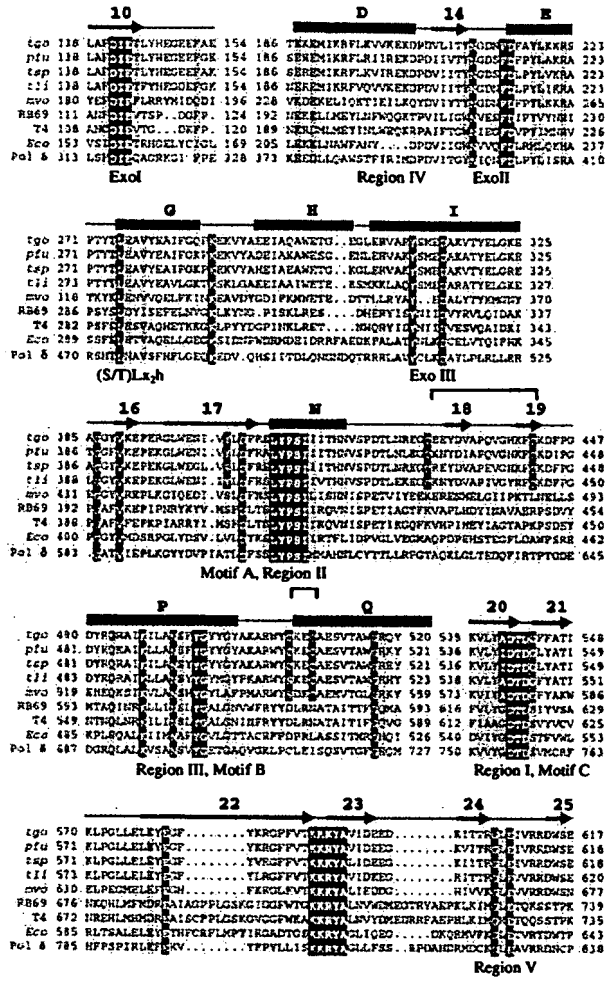


FIG. 3. Sequence alignment of B family DNA polymerases. The alignment has been adapted from ref. 21 to highlight specific residues from the class of archaeal pols. The secondary structure of *Tgo* pol is indicated on top of the alignment with helices (bars), strands (arrows) and loops (lines) colored according to domains with the same color code as Fig. 2*B*. Strictly conserved residues of type B polymerases are red, and additional conserved residues are yellow. Uniquely conserved residues of archaeal type B enzymes—as discussed in the text—are green. Disulfide bonds are shown by a bar on top of the alignment. Abbreviations: *tgo*, *Thermococcus gorgonarius* pol; *pfu*, *Pyrococcus furiosus* pol; *tsp*, *Thermococcus* sp. pol; *tli*, *Thermococcus litoralis* pol; *mvo*, *Methanococcus voltae* pol; RB69, bacteriophage RB69 pol; T4, bacteriophage T4 pol; *Eco*; *E. coli* pol II; pol δ; human pol δ.

packed against the five-stranded antiparallel β-sheet that contains the three conserved aspartate residues involved in nucleotidyl transfer. The fingers emerge from the palm domain as an α-helix-rich insertion. Its 50 residues are folded into two antiparallel coiled α-helices of approximately equal size: helix P contains the conserved $Kx_3NsxYGx_2G$ motif of B type polymerases and is related to the O helix of A type enzymes (see below). The ≈50-residue insertion between helix O and P in RB69 and T4 gp43 is missing in *Tgo* pol, where both helices and 4 residue linker are much shorter than their equivalents in gp43. The shorter fingers of *Tgo* pol presumably reflect the typical structure of the nonbacteriophage B type fingers (pol α, pol δ, and *E. coli* pol II). The thumb domain topology, similar to that of gp69, is unrelated to other polymerase types. However, like the thumb of A type enzymes, a bundle of α-helices at its base protrude from the active site β-sheet. Distal to the active site, the thumb contains a 75-residue subdomain (665–729), which fixes the exonuclease domain and contributes to the editing channel, explaining why mutations in the exonuclease domain of B-type polymerases affect the polymerase activity and vice versa (44; 45).

Weakly defined density across the base of the thumb domain was modeled as the C-terminal 6 residues with a polyalanine chain. The C terminus thus does not protrude from the core molecule as in the RB69 polymerase (21). Because the C terminus of the T4 pol are involved in sliding-clamp binding (46), it is likely, however, that these residues become ordered on any similar holoenzyme formation.

Sequence Alignment of Archaeal DNA Polymerases. The structure of *Tgo* pol allows the generation of a structure based sequence alignment of the archaeal subfamily of type B DNA polymerases, the location of conserved and unique residues, and the comparison with other type B DNA polymerases (Fig. 3).

Polymerase Active Site. The polymerase active site is formed by the central β-sheet (strands 16, 17, 20, 21, and 22) and helix N of the palm domain and helix P located in the fingers and is highly conserved among B family polymerases (Fig. 4). Three carboxylates required for nucleotidyl transfer in B family polymerases, two of which coordinate two metal ions (14) are superimposably conserved among A family enzymes, B family enzymes, and reverse transcriptase (21). Superposition of *Tgo* pol and T7 replication complex (14) places the dNTP near the proposed nucleotide-binding site in helix P, the $K487x_3NSxYGx_2G$ motif (Fig. 3) and suggests interactions of the carboxylates with metals and the phosphate tail of the bound dNTP (Fig. 4). Reorientation of the strictly conserved Lys-487 allows it to mimic the Lys-522–phosphate tail interaction in T7. Tyr-494 ($Kx_3NSxY494Gx_2G$) and Tyr-409 (SLY409PSII) form the bottom of the nucleotide-binding site.

The active site of B family polymerases contains a DTDS motif, which, however is DTDG in the archaeal subfamily. In *Tgo* pol, the relatively conserved Tyr-402 from the adjacent strand provides an alternate alcohol group, at a position appropriate for metal coordination or binding of the 3′ end of the primer. The orientation of Tyr-402 is stabilized by an aromatic cluster that also includes Phe-545 and Tyr-538. Archaeal *Methanococcus voltae* and *Thermococcus* sp. pol's (see Fig. 3) have Tyr at position 545—Phe in *Tgo*—rather than at 402, but might also supply an alcohol group. The displacement of a functional alcohol from serine in DTDS to Tyr-402 or Tyr-545 might stabilize its orientation as an adaptation for thermostability.

The conserved cluster of acidic amino acids (E578, E580) form an unexpected metal-binding site for Mn^{2+} and Zn^{2+} (Fig. 4). Its proximity to Asp-404 and to the expected location of the dNTP γ-phosphate suggests a supporting role in nucleotide binding and/or catalysis.

3′ → 5′ Exonuclease Active Site. *Tgo* pol is characterized by a strong 3′ → 5′ exonuclease activity, unlike eukaryotic B type polymerases (unpublished results). The exonuclease active site is formed at the interface between the exonuclease domain and the tip of the thumb (Fig. 5). All residues required for catalysis are located in the exonuclease domain, which, at least for T4 gp43, retains activity when dissociated from the polymerase (47). However, the thumb domain, with, for example, RB69 gp43's Phe-123–base intercalation, partially controls the binding geometry of single strand DNA (21, 43).

The exonuclease structures of *Tgo* and gp43 DNA polymerases are similar at the editing site but differ considerably at the exonuclease–thumb interface. Strand 10 contains the metal-binding D141IE motif and readily superimposes with the equivalent strand from gp43, allowing modeling of a single-strand DNA segment into the exonuclease site based on theRB69 gp43-$p(dT)_4$ complexes (21). The conserved residues Asp-141 and Glu-143 in the Exo I motif, Tyr-209, Asn-210, Phe-214, and Asp-215 in Exo II, and Tyr-311 and Asp-315 in Exo III are in approximate DNA-binding conformations (Fig.

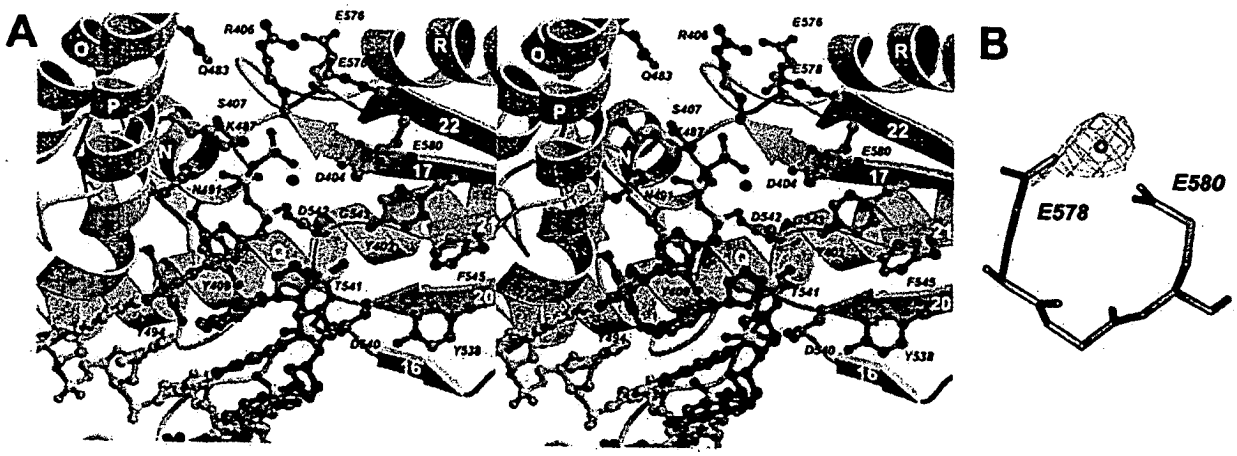


FIG. 4. Polymerase active site. (*A*) Stereoribbon representation (using color code as in Fig. 2) with modeled DNA. Active-site residues are shown as ball-and-stick representations with carbon (green), nitrogen (blue), and oxygen (red) atoms. The DNA template (light brown), primer (light brown), and dNTP (orange) complex has been taken from the coordinates of T7 replication complex (15) by superimposing D404, D542, and adjacent residues with corresponding residues in T7 pol (D475 and D654). Phosphorus atoms are yellow. The two metals of the T7 replication complex are shown as magenta spheres. (*B*) Experimentally observed metal-binding site for Mn^{2+} ($F_o - F_c$ density contoured at 5σ) and Zn^{2+} in the "low salt" crystal form. The carboxylates E578 and E580 are conserved in type B polymerases (Fig. 3).

5*A*). However, the editing cleft is constricted by a displacement of the tip of the thumb toward the exonuclease domain to prohibit single-strand binding (Fig. 5*B*). This shift is correlated with a large change in the loop between strands 10 and 11. In RB69 gp43 (and likewise in T4 gp43), this loop forms a lid over the 3′ base and contains Phe-123, which intercalates between the first two bases. In *Tgo* pol, the loop is curved outward, away from the thumb, and Phe-152 (the equivalent of gp43's Phe-123) attaches to Phe-214 10 Å away from intercalation site. This shift allows the tip of the thumb to move into the editing channel and to block the exonuclease site.

Are There Different Conformations in Polymerase and Editing Mode? If a closed conformation of the exonuclease domain prohibits single strand binding, an open conformation is required for editing. The observed closed conformation may represents the enzyme in "polymerase" mode. Preliminary analysis of the crystal structure of *Tgo* pol in the low-salt conditions indicates a structural change at the interface of exonuclease and thumb, possibly reflecting a transition between open and closed forms. The closed conformation observed here may, however, be a nonphysiological artifact of the high ionic strength used for crystallization. Crystal structures of the enzyme in both polymerase and editing modes are required.

Adaptation to High Temperatures. *Tgo* is a sulfur-metabolizing, extremely thermophilic archaeon, with a growth range between 55°C and 98°C. For accurate replication at this temperature range, the polymerase must not only be stable, but must also adequately bind substrate DNA. A comparison with gp43 from the mesophilic bacteriophage RB69 indicates several such adaptations to high temperatures. Several loops are shorter in *Tgo* pol than in gp43 (Fig. 2), and there is an increase in hydrogen bonded β-strand content: *Tgo* pol secondary structure includes 41% helix, 22% β-strands, and 19% turns (calculated according to ref. 48), whereas gp43 has 42% helix, 17% β-strands, and 19% turns.

Although rare among cytoplasmic or nuclear proteins, two disulfide bridges might be formed: cysteine pairs 428–442 and

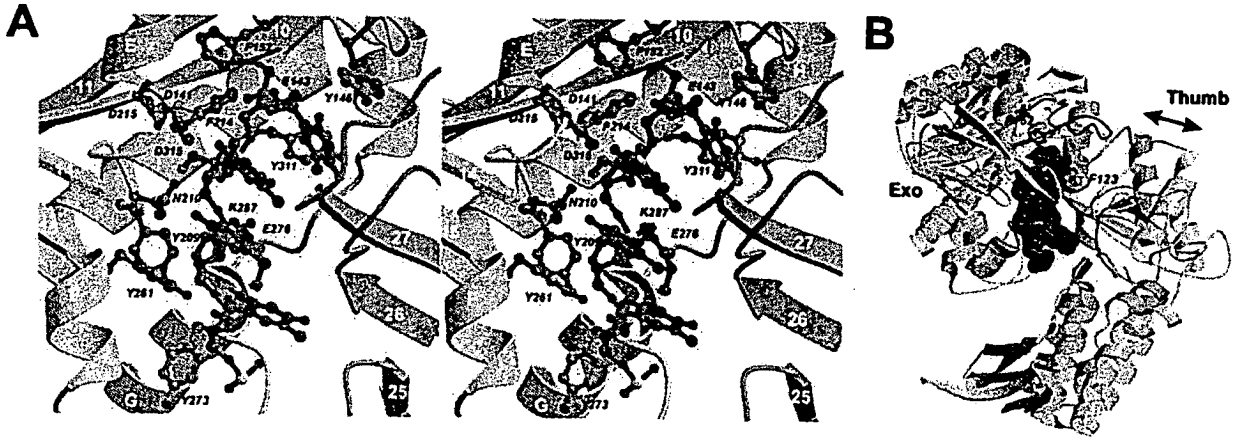


FIG. 5. 3′ → 5′ exonuclease active site. (*A*) Stereoribbon representation with modeled DNA using the color code of Fig. 4*A* (ball-and-stick) and Fig. 2 (ribbons). Active-site residues are shown as ball-and-stick representation. The single-stranded DNA has been taken from the coordinates of RB69–single-stranded DNA complex (21). The orientation of the DNA has been obtained by superimposing the DIE143 motif in *Tgo* pol with corresponding motif of RB69 gp43 (DIE116). Strand 27 and its preceding loop from the thumb (green) is apparently in collision with the modeled DNA. (*B*) Comparison of the exonuclease–thumb interface between *Tgo* pol (color code of Fig. 2*B*) and RB69 gp43 (gray). In *Tgo* pol, the lid of the editing site (red) is bent outward compared with the equivalent loop of gp43 (yellow), allowing the tip of the thumb to move several Å (↖↘) closer to the exonuclease. This conformation is incompatible with formation of an editing complex [the $p(dT)_4$ of gp43 is shown as brown space-filling model].

506–509, although reduced, are poised for attachment (Fig. 2). Model refinement and electron density inspection with and without constraints for the disulfide bonds verified the reduced state (observed unrestrained SG–SG distance: 2.8 Å and 3.0 Å). This is consistent with our *E. coli* expression and further rules out structural perturbation by nonnative oxidation. These cysteines are located in the palm domain and are conserved among B type enzymes from hyperthermophilic sulfur-metabolizing archaeons, but not among mesophile homologs (Fig. 3). The Cys-428–Cys-442 bridge stabilizes the compact fold of the loop segment between helix N in the palm domain and helix O in the fingers and presumably also the relative orientation of these helices. In addition, the loop segment packs against helix Q in the palm domain. Helix Q, the spine of the palm domain, is further stabilized at the first helical turn by the second disulfide bridge between Cys-506 and Cys-509.

A much enhanced complementary positive potential for all three DNA-binding clefts of *Tgo* pol is observed relative to gp43 (Fig. 2). Thus, in addition to hydrogen bonding and specific DNA–protein interactions, binding to *Tgo* pol has an additional strong stabilizing electrostatic component.

An increase in the number of salt bridges is often associated with thermostability. Although *Tgo* pol has a greater total number of charged residues (262) than gp43 (245), both molecules have 54 salt bridges within a 3–5 Å bound. However, in the 5–7 Å range of charge distance, *Tgo* pol has 77 ion pairs compared with 43 for gp43. This large increase results in a more highly charged surface of *Tgo* pol, accompanied by a more balanced charge distribution, compared with gp43 where charges are often located in patches (Fig. 2).

We thank Dr. Uwe Jacob and Martin Augustin for stimulating discussions, Dr. Nediljko Budisa for help with expression of selenomethionine variants, and Dr. Hans Bartunik and Dr. Gleb Bourenkov at DESY, Hamburg for help with data collection.

1. Waga, S. & Stillman, B. (1998) *Annu. Rev. Biochem.* **67**, 721–751.
2. Maga, G. & Hübscher, U. (1996) *Biochemistry* **35**, 5764–5777.
3. Wang, T. S. F. (1991) *Annu. Rev. Biochem.* **60**, 513–552.
4. Kornberg, A. & Baker, T. A. (1992) *DNA Replication* (Freeman, New York).
5. Joyce, C. M. & Steitz, T. A. (1994) *Annu. Rev. Biochem.* **63**, 777–822.
6. Perler, F. B., Kumar, S. & Kong, H. (1996) *Adv. Protein Chem.* **48**, 377–435.
7. Brautigam, C. A. & Steitz, T. A. (1998) *Curr. Opin. Struct. Biol.* **8**, 54–63.
8. Arnold, E., Ding, J., Hughes, S. H. & Hostomsky, Z. (1995) *Curr. Opin. Struct. Biol.* **5**, 27–38.
9. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) *Nature (London)* **313**, 762–766.
10. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. & Steitz, T. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8924–8928.
11. Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. & Beese, L. S. (1997) *Structure (London)* **5**, 95–108.
12. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998) *Nature (London)* **391**, 304–307.
13. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1995) *Nature (London)* **376**, 612–616.
14. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) *Nature (London)* **391**, 251–258.
15. Davies, J. F., Almassy, R. J., Hostomska, Z., Ferre, R. A. & Hostomsky, Z. (1994) *Cell* **76**, 1123–1133.
16. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H. & Kraut, J. (1994) *Science* **264**, 1930–1935.
17. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) *Science* **264**, 1891–1903.
18. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992) *Science* **256**, 1783–1790.
19. Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark P., *et al.* (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6320–6324.
20. Jager, J., Smerdon, S. J., Wang, J., Boisvert, D. C. & Steitz, T. A. (1994) *Structure (London)* **2**, 869–876.
21. Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997) *Cell* **89**, 1087–1099.
22. Edgell, D. R. & Doolittle, W. F. (1997) *Cell* **89**, 995–998.
23. Cann, I. K. O., Komori, K., Toh, H., Kanai, S. & Ishino, Y. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14250–14255.
24. Debyser, Z., Tabor, S. & Richardson, C. C. (1994) *Cell* **77**, 157–166.
25. Nossal, N. G. (1994) *Molecular Biology of Bacteriophage T4*, ed. Karam, J. D. (Am. Soc. Microbiol., Washington, DC), pp. 43–53.
26. Nossal, N. G. (1992) *FASEB J.* **6**(3), 871–878.
27. Fanning, E. & Knippers, R. (1992) *Annu. Rev. Biochem.* **61**, 55–85.
28. Kelman, Z. & O'Donnell, M. (1996) *Annu. Rev. Biochem.* **64**, 171–200.
29. De Pamphilis, M. L., ed. (1996) *DNA Replication in Eukaryotic Cells* (Cold Spring Harbor Lab. Press, Plainview, NY).
30. Waga, S. & Stillman, B. (1994) *Nature (London)* **369**, 207–212.
31. Baker, T. A. & Bell, S. P. (1998) *Cell* **92**, 295–305.
32. Uemori, T., Sato, Y., Kato, I., Doi, H. & Ishino, Y. (1997) *Genes Cells* **2**(8), 499–512.
33. Braithwaite, D. K. & Ito, J. (1993) *Nucleic Acids Res.* **21**(4), 787–802.
34. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayre, J. D. & Venter, J. C. (1996) *Science* **273**, 1058–1073.
35. Miroshnichenko, M. L., Gongadze, G. M., Rainey, F. A., Kostyukova A. S., Lysenko, A. M., Chernyh, N. A. & Bonch-Osmolovskaya, E. A. (1998) *Int. J. Syst. Bacteriol.* **48**, 23–29.
36. Bonch-Osmolovskaya, E., Svetlichny, V., Ankenbauer, W., Schmitz-Agheguian, G., Angerer, B., Ebenbichler, C. & Laue, F. (1996) Eur. Patent Appl. EP 0 834 570 A1.
37. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995) *Eur. J. Biochem.* **320**, 788–796.
38. Leslie, A. G. W. (1994) MOSFLM User Guide (MRC Laboratory of Molecular Biology, Cambridge, UK), Version 5.1.
39. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307–326.
40. Collaborative Computational Project, No. 4 (1994) *Acta Crystallogr. D* **50**, 760–763.
41. Turk, D. (1992) Ph.D. Thesis (Technische Universität, München).
42. Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 5018–5023.
43. Wang, J., Yu, P., Lin, T. C. Konigsberg, W. H. & Steitz, T. A. (1996) *Biochemistry* **35**, 8110–8119.
44. Sattar A. K., Lin, T. C., Jones, C. & Konigsberg, W. H. (1996) *Biochemistry* **35**, 16621–16629.
45. Reha-Krantz, L. J. & Nonay, R. L. (1994) *J. Biol. Chem.* **269**, 5635–5643.
46. Berdis, A. J., Soumillion, P. & Benkovic, S. J. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12822–12827.
47. Lin, T. C., Karam, G. & Konigsberg, W. H. (1994) *J. Biol. Chem.* **269**, 19286–19394.
48. Kabsch, W. & Sander C. (1983) *Biopolymers* **22**, 2577–2637.

Crystal structure of an archaebacterial DNA polymerase

Yanxiang Zhao[1†], David Jeruzalmi[1†], Ismail Moarefi[1,2], Lore Leighton[1,2], Roger Lasken[3] and John Kuriyan[1,2*]

Background: Members of the Pol II family of DNA polymerases are responsible for chromosomal replication in eukaryotes, and carry out highly processive DNA replication when attached to ring-shaped processivity clamps. The sequences of Pol II polymerases are distinct from those of members of the well-studied Pol I family of DNA polymerases. The DNA polymerase from the archaebacterium *Desulfurococcus* strain Tok (D. Tok Pol) is a member of the Pol II family that retains catalytic activity at elevated temperatures.

Results: The crystal structure of D. Tok Pol has been determined at 2.4 Å resolution. The architecture of this Pol II type DNA polymerase resembles that of the DNA polymerase from the bacteriophage RB69, with which it shares less than ~20% sequence identity. As in RB69, the central catalytic region of the DNA polymerase is located within the 'palm' subdomain and is strikingly similar in structure to the corresponding regions of Pol I type DNA polymerases. The structural scaffold that surrounds the catalytic core in D. Tok Pol is unrelated in structure to that of Pol I type polymerases. The 3′–5′ proofreading exonuclease domain of D. Tok Pol resembles the corresponding domains of RB69 Pol and Pol I type DNA polymerases. The exonuclease domain in D. Tok Pol is located in the same position relative to the polymerase domain as seen in RB69, and on the opposite side of the palm subdomain compared to its location in Pol I type polymerases. The N-terminal domain of D. Tok Pol has structural similarity to RNA-binding domains. Sequence alignments suggest that this domain is conserved in the eukaryotic DNA polymerases δ and ε.

Conclusions: The structure of D. Tok Pol confirms that the modes of binding of the template and extrusion of newly synthesized duplex DNA are likely to be similar in both Pol II and Pol I type DNA polymerases. However, the mechanism by which the newly synthesized product transits in and out of the proofreading exonuclease domain has to be quite different. The discovery of a domain that seems to be an RNA-binding module raises the possibility that Pol II family members interact with RNA.

Addresses: [1]Laboratories of Molecular Biophysics, [2]Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA and [3]Molecular Staging, Inc., 66 High Street, Guilford, CT 06437, USA.

[†]The first two authors contributed equally to this work.

*Corresponding author.
E-mail: kuriyan@mail.rockefeller.edu

Key words: *Desulfurococcus* Tok, DNA polymerase, Pol II family, RNA-binding domain, thermostability, X-ray crystallography

Received: **12 April 1999**
Revisions requested: **27 May 1999**
Revisions received: **9 June 1999**
Accepted: **9 June 1999**

Published: **20 September 1999**

Structure October 1999, 7:1189–1199
http://biomednet.com/elecref/0969212600701189

0969-2126/99/$ – see front matter

Introduction

DNA polymerases can be classified into at least three families on the basis of sequence similarities to the three distinct DNA polymerases of *Escherichia coli*, Pol I, Pol II and Pol III [1]. Members of the Pol I family have been studied extensively, resulting in a comprehensive understanding of their functional properties and their structure [2–6]. In contrast to the detailed knowledge that is now available for the Pol I family, the Pol II and Pol III polymerases are poorly understood. The first crystal structure determined for a Pol II family member was that of the DNA polymerase of the bacteriophage RB69 (RB69 Pol) [7] and no structural information is currently available for any member of the Pol III family. Members of the Pol II (also known as Pol B or Pol α) and Pol III families carry out processive replication of chromosomal DNA during cell division [8], and there is interest in further extending our knowledge of their structures and mechanism. Archaebacterial DNA polymerases and the eukaryotic DNA polymerases α, δ and ε are members of the Pol II family [1].

The structure of RB69 Pol revealed that the general architecture of the core of the Pol II polymerases is strikingly similar to that of the Pol I polymerases [7]. Pol I polymerases are constructed from three smaller subdomains, termed the thumb, palm and fingers regions by analogy to elements first noted in the structure of the Klenow fragment of *E. coli* DNA polymerase I [9]. In addition, Pol I DNA polymerases have a proofreading 3′–5′ exonuclease domain located below the thumb subdomain, near the region where duplex DNA exits the polymerase active site [4,5]. Besides the residues involved in catalysis, there is no significant sequence similarity between the polymerase domains of members of the Pol I and Pol II families [1].

However, the subdomain architecture of the Pol I family is conserved in the RB69 structure, even though the detailed structures of the subdomains are quite divergent [7]. The exonuclease domains of Pol I and Pol II DNA polymerases are closely related in sequence and, not surprisingly, the structure of the exonuclease domain of RB69 resembles that of the Pol I type polymerases. Given the general similarity in the polymerase domains of the Pol I polymerases and RB69, the location of the exonuclease domain in RB69 was a surprise. In RB69 the 3′–5′ exonuclease domain is located above the fingers and opposite the thumb subdomains, suggesting that the shuttling of DNA between the polymerization and proofreading sites must occur by a different mechanism in Pol II DNA polymerases [7].

The mechanism of the Pol I family DNA polymerases is now understood in detail [4,5,10,11,28]. The chemistry of nucleotide addition is mediated by two metal ions that are liganded by two aspartate residues. These are located in the palm subdomain, at the base of a deep cleft in the polymerase domain. High-resolution crystal structures of the Pol I type DNA polymerases of T7 bacteriophage (T7 Pol) and *Thermus aquaticus* (Taq Pol) complexed to primer-template DNA and incoming nucleotide have been determined, allowing the mechanisms of nucleotide incorporation and selectivity to be visualized [10,11,28]. Although corresponding structural information for the Pol II family DNA polymerases is lacking, similarities in general organization of the polymerase core as well as sequence conservation within crucial elements of the central palm subdomain suggest that general features of the recognition of DNA will be similar in Pol II polymerases.

The DNA polymerase from the archaebacterium *Desulfurococcus* strain Tok (D. Tok Pol) is a member of the Pol II family, and has both thermostable DNA polymerase and 3′–5′ exonuclease activities [12]. D. Tok Pol sustains undiminished DNA polymerase activity after incubation at 95°C for one hour (RL, unpublished results). The sequence of D. Tok Pol is very closely related (> 75% identity) to that of other archaebacterial DNA polymerases, such as those from *Pyrococcus furiosus* [13] and *Thermococcus littoralis* [14]. D. Tok Pol is also related to eukaryotic DNA polymerases α, δ and ε (34% sequence identity over 196 residues of the DNA polymerase core for the human δ sequences) [1]. The archaebacterial genomes also contain genes coding for proteins with clear homology to proliferating cell nuclear antigen (PCNA), the DNA polymerase clamp in eukaryotes, as well as subunits of the clamp-loader complex RF-C (replication factor C). It is likely that archaebacterial DNA polymerases achieve processivity by attachment to the ring-shaped PCNA ring, although direct evidence for such a mechanism is lacking.

We have determined the structure of D. Tok Pol at 2.4 Å resolution. D. Tok Pol shares less than 20% sequence identity with RB69 Pol, but the structures of the two enzymes resemble each other closely. The structure reported here has been determined in the absence of DNA. Nevertheless, the close structural correspondence between the active sites of Pol I and Pol II DNA polymerases allows inferences to be made about the mode of DNA recognition by D. Tok Pol. The very N-terminal region of D. Tok Pol contains a domain (residues 1–132) that is closely related in structure to single-stranded RNA-binding domains (RBDs), also known as RNA-recognition modules (RRMs) [15]. The structure of the 3′–5′ proofreading exonuclease domain of D. Tok Pol is similar to those of the Pol I type polymerases. However, its location relative to the palm subdomain resembles the location seen in RB69 [7] rather than the Pol I type polymerases [9,16,23]. The structure of D. Tok Pol reported here provides further evidence that the mode of DNA-template recognition and the distinct editing channel established for the Pol II family by the structure of RB69 Pol is valid for the entire Pol II family.

Results and discussion

Structure determination

Crystals of D. Tok Pol have been obtained from 2,4-methylpentanediol (MPD) (Native I) and polyethylene glycol (PEG) 400 (Native II). Both crystal forms are orthorhombic ($P2_12_12_1$; a = 64.8 Å, b = 107.6 Å, c = 153.2 Å for Native I and a = 66.1 Å, b = 107.6 Å, c = 155.9 Å for Native II). Experimental phases (Table 1) to 3.0 Å were obtained from four isomorphous heavy-atom derivatives, using Native II and the program SHARP [17]. Phases were improved by iterative cycles of real-space density modification, consisting of solvent flipping and negative density truncation, using SOLOMON [18,19]. The resulting electron-density map allowed the chain to be traced unambiguously, with ready determination of sequence register. The model was refined to 2.6 Å against data for Native II (R value = 24.2%, R_{free} = 29.5%) and subsequently to 2.4 Å against data for Native I (R value = 25.3%, R_{free} = 29.9%), using CNS [20]. The model for Native II is somewhat more complete (see the Materials and methods section) and is used for most of the discussion. This model includes 740 residues from 1 to 756 in Native II. Amino acids 386–390 and 665–676 are not visible in our electron-density maps and are not included in the model.

General description of the structure

D. Tok Pol (Figure 1) is composed of a polymerase domain (residues 390–773) and an exonuclease domain (residues 133–385), as well as an N-terminal domain (residues 1–131) that is not found in Pol I type DNA polymerases [4]. The polymerase domain is further comprised of three smaller subdomains, termed the thumb (residues 607–756), palm (residues 390–445 and 500–606) and fingers (residues 446–499). The structures of the MPD and PEG400 crystal forms of D. Tok Pol are very similar

Table 1

Data collection, structure determination and refinement statistics.

	Resolution (Å)	Number of reflections (unique)	Completeness (%)	R_{sym}* (%)	R_{iso}† (%)	Sites	Phasing power‡	Figure of merit§
Native data	–	–	–	–	–	–	–	–
Native II	50.0–2.6	32,909	93.9(57.8)	5.2(15.6)	–	–	–	–
Native I	50.0–2.4	40,540	92.2(53.8)	4.6(31.9)	57.4			
MIRAS analysis	–	–	–	–	–	–	–	0.367
Pt	50.0–3.0	40,316	97.9(93.0)	8.4(22.9)	19.1	4	1.34(0.99)	0.214
Pb	50.0–3.0	34,905	84.2(70.4)	5.9(18.6)	13.6	1	1.16(0.98)	0.195
Pt/Pb	50.0–3.0	35,107	80.8(57.4)	9.9(21.9)	18.3	5	1.59(0.80)	0.221

Refinement	Number of Reflections (|F| > 2σ)		$R_{working}$#/R_{free}¶ (%)	Total number of atoms	Rmsd for bonds (Å)	Rmsd for angles (°)	Rmsd for B values (Å²)
Native II	50.0–2.6	31,591	24.2/29.5	6,167	0.008273	1.61591	1.691
Native I	50.0–2.4	37,229	25.3/29.9	6,145	0.008273	1.50479	1.409

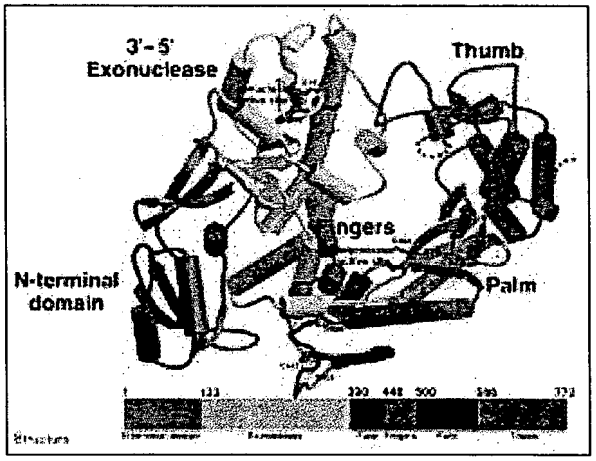
*R_{sym}% = 100 X Σ|I – < I >| / ΣI, where I is the integrated intensity of a given reflection. †R_{iso}% = 100 X Σ|F_{PH}–F_P| / ΣF_P, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively. ‡Phasing power = Σ|F_H| / Σ||F_{PH}(obs)| – |F_{PH}(calc)||, where F_H is the calculated heavy atom structure factor amplitude. §Figure of merit = <|ΣP(α)$e^{iα}$/Σ|P(α)|>, where α is the phase and P(α) is the phase probability distribution. #Rworking = Σ|F(obs) – F (calc)| / ΣF(obs). ¶R_{free} = Σ|F(obs) – F (calc)| / ΣF(obs), calculated using 10% of the data. Numbers in parentheses apply to the highest resolution shell.

in terms of the individual subunits. The major difference between the two structures is a rotation of ~8–10° in the orientation of the exonuclease domain with respect to the thumb subdomain.

The domains of D. Tok Pol are arranged as an irregularly shaped flattened ring with a central cavity located near the polymerase active site. The mostly α-helical thumb subdomain forms one side of the active-site cleft and makes contacts with the exonuclease domain (Figure 1). The structures of the thumb domains of various polymerases are often unrelated in structure. However, in all cases where structures are available the thumb domain is seen to fulfil an important role by forming contacts with duplex DNA as it exits the polymerase active site [4]. The D. Tok Pol structure has been determined in the absence of DNA, and a portion of the thumb subdomain that is likely to contact DNA (residues 665–676) is disordered. This is commonly observed for the corresponding regions of other polymerases in the absence of substrate [9,21–24]. In the DNA polymerases from bacteriophage T4 and RB69, the thumb subdomains also provide a C-terminal element that interacts with the processivity clamp [25,26]. In D. Tok Pol, the corresponding region (residues 757–773) is disordered.

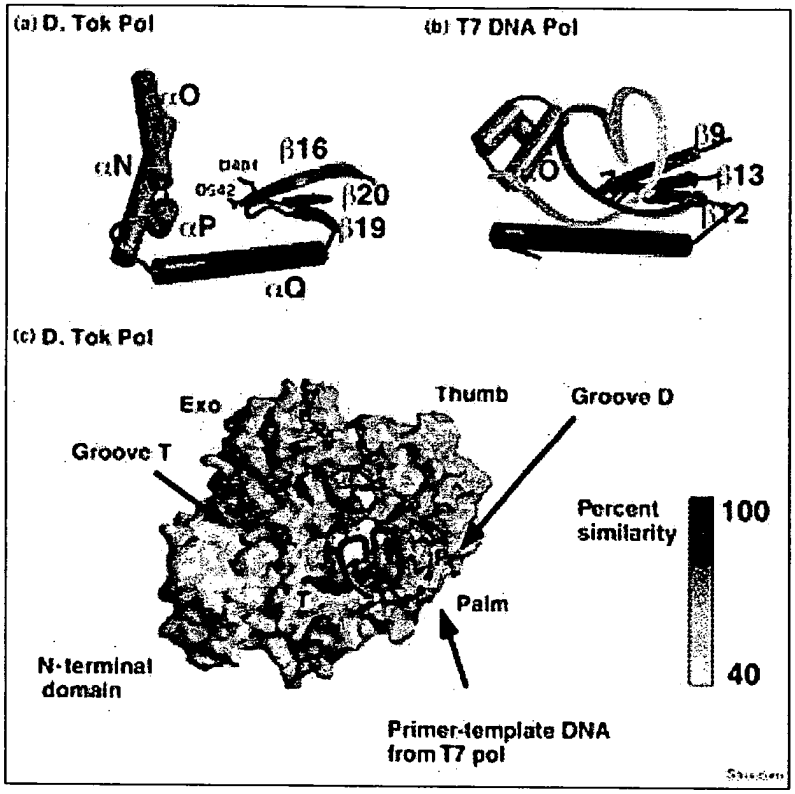
The central region of the active-site cleft is occupied by the palm subdomain and includes residues important for substrate discrimination and the catalysis of the polymerase reaction. In D. Tok Pol, the palm is organized around three β strands (β16, β19, β20) flanked by an α helix (αQ) (Figures 1,2a,3a). It contains two disulfide bonds (Cys428–Cys442, Cys506–Cys509) that have not been previously observed in palm subdomains and which may be important for thermostability (Figure 1).

Figure 1



Structure of D. Tok Pol. The structure is represented by cylinders for helices, arrows for strands, and a thin worm for other secondary structural elements. Two gray spheres represent metal ions (presumed to be Mg^{2+}) observed to be bound to the exonuclease domain. The active site of the polymerase is marked by the location of two aspartate residues D404 and D542. The two disulfide bonds are indicated. Regions of the polypeptide chain that could not be modeled in the palm subdomain because of disorder are indicated by dotted lines. The various domains and subdomains and their boundaries are indicated in the bar.

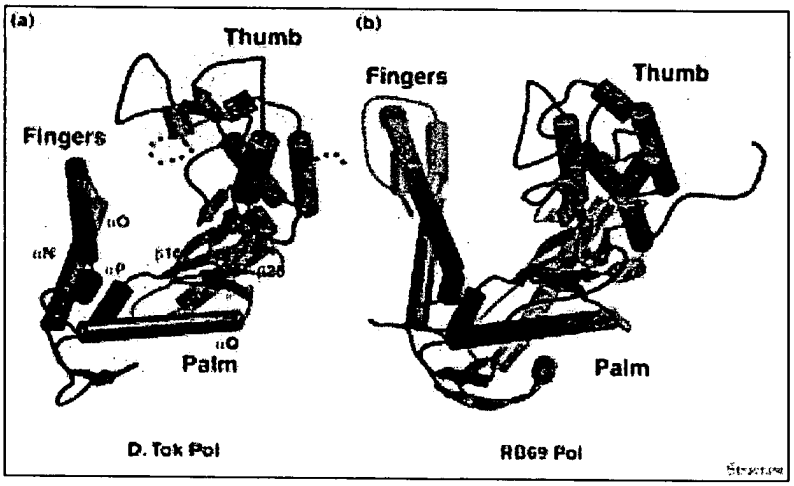
Figure 2



Comparison of DNA polymerase structures. **(a)** A view of the secondary structural elements of the polymerase active-site region (palm and fingers subdomains) of D. Tok Pol, colored as in Figure 1. **(b)** The corresponding region of T7 DNA polymerase including the primer-template duplex from the crystal structure (PDB code 1T7P [11]). The orientation of T7 Pol was derived by superposition onto strands β16, β19, and β20 of D. Tok Pol. D. Tok Pol helix αP is seen to be in an analogous position relative to the active-site aspartates as T7 Pol αO. **(c)** A GRASP surface representation of D. Tok Pol with modeled primer-template duplex from the T7 DNA polymerase–DNA complex (PDB code 1T7P [11]). The surface is colored according to sequence similarity (40–100%) calculated as in Figure 7c. The primer strand is an orange worm representing phosphate positions, and the template strand is in gray.

The central elements of the palm subdomains from polymerases belonging to the Pol I and Pol II families can be aligned closely (the root mean square deviation [rmsd] in Cα positions for strands β16, β19, β20 and helix αQ is in the range of 0.9–2.0 Å), indicating a potential conservation of function. There are two residues in the palm domains

Figure 3



Comparison of the structures of D. Tok Pol and RB69 Pol. The structures of **(a)** D. Tok Pol and **(b)** RB69 Pol are presented in the same orientation after superposition of their respective palm subdomains. Structural elements that are in common between the structures are represented and colored as in Figure 1. Elements that are unique to RB69 Pol are colored in gray. Disordered segments are indicated by dotted lines. The N-terminal and the exonuclease domains are not shown.

of Pol I polymerases that are crucial for enzymatic activity because they coordinate two metal ions [2,10,11,27]. The corresponding residues in D. Tok Pol are Asp404 and Asp542 (Figure 1). No metal ions are, however, visible in our electron-density maps.

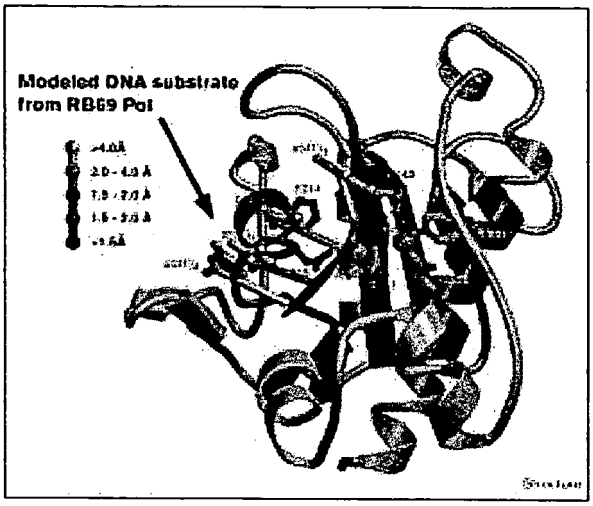
The fingers subdomain in D. Tok Pol consists of a set of antiparallel α helices (αN, αO, αP; Figure 2). These helices are shorter in length than the corresponding elements of RB69 Pol, and a helical segment that connects helices O and N in RB69 Pol is missing altogether (Figure 3). The fingers domain of D. Tok Pol is unrelated in overall structure to that of Pol I type polymerases (Figure 2). However, helix αP in D. Tok Pol is positioned similarly to helix O in Pol I polymerases (Figure 2), and is likely to play an analogous and crucial role in recognition of the incoming nucleotide [9–11,28].

The 3′–5′ exonuclease domain in D. Tok Pol is located opposite the thumb subdomain and above the fingers subdomain, as noted for RB69 Pol. It contains two metal ions (presumably Mg^{2+}) ligated to Asp141 and Glu143 (Figure 1). The position of this domain relative to the polymerase active site is distinct from the arrangement seen in Pol I type polymerases. The conservation between RB69 and D. Tok Pol of the location of the exonuclease domain suggests that this is a characteristic feature of Pol II type polymerases. The structure of the D. Tok Pol 3′–5′ exonuclease domain resembles those associated with other DNA polymerases [29,30]. The 3′–5′ exonuclease domains from the Pol I (*E. coli*, *T. aquaticus*, *Bacillus subtilis*, bacteriophage T7) or Pol II (RB69) polymerase families can be aligned onto each other closely (rmsd in Cα positions for strands β10, β11, β12, β14 and helices αE and αI is in the range of 1.0–2.8 Å). This alignment superimposes residues associated with substrate binding, catalysis and metal binding in a satisfactory manner (Figure 4) [4].

The arrangement of the N-terminal, exonuclease, and polymerase domains creates two deep grooves leading into and out of the polymerase active site. The D groove (for duplex-DNA binding, following the nomenclature of [7]) is located immediately below the thumb subdomain and includes a region of positive electrostatic potential. The T groove (for template-DNA binding) leads away from the active site in the opposite direction and is located below the fingers subdomain. A small channel (the editing channel) leads from the polymerase domain to the exonuclease active site (Figures 2c).

We have used the structure of T7 Pol bound to primer-template DNA to model DNA onto D. Tok Pol (Figure 2c). Superposition of the palm subdomains of the two polymerases shows that remarkably few bad contacts are formed between the DNA (from T7 Pol) and atoms in the D. Tok Pol model. The one region that does collide with the DNA is the segment connecting the exonuclease and polymerase domains. This region (residues 377–390) is partially disordered in the D. Tok Pol structures, and is likely to reorganize upon binding DNA. This superposition allows five base pairs of DNA to be accommodated in the D. Tok Pol active site, with the formation of DNA–protein contacts. The formation of contacts with additional base pairs would require a change in the position of the thumb subdomain in the region of the D groove. A change in the conformation of the fingers subdomain (helices αO and αP) is also required to position residues Lys487 and Tyr493 (or Tyr494) of D. Tok Pol (Figure 2) for interaction with the incoming nucleotide, by analogy with the T7 Pol structure [11]. Finally, the superimposed primer-template DNA is well positioned so that the incoming template strand will probably reside in the T groove. Superposition of the DNA molecule derived from the structure of HIV-1 reverse transcriptase complexed to DNA [31] leads to similar conclusions.

Figure 4



Structural alignment of exonuclease domains. Structures of exonuclease domains from KF, 1WAJ, 1T7P, 1BDF, and 1TAQ have been aligned by superimposing residues 137–145, 158–164, 167–172, 205–220, 257–260, and 303–313, which represent strands β8, β9, β10, β12, β15 and helix αE, αI. A color gradient is used to depict the average rmsd for the family of superimposed structures ranging from blue (1.0–1.5 Å) to white (> 4.0 Å). Residues conserved amongst exonuclease sequences and implicated in catalysis are drawn in green ball and stick representation. Two gray spheres represent two metal ions bound at the active site. The active site is also indicated by a tetranucleotide (in gold) derived from superposition of the exonuclease domain from the RB69 Pol structure.

Comparison between D. Tok Pol and RB69 Pol

Although the DNA polymerases from D. Tok Pol and bacteriophage RB69 share less that 20% primary sequence identity (Figure 5), their structures resemble each other

Figure 5

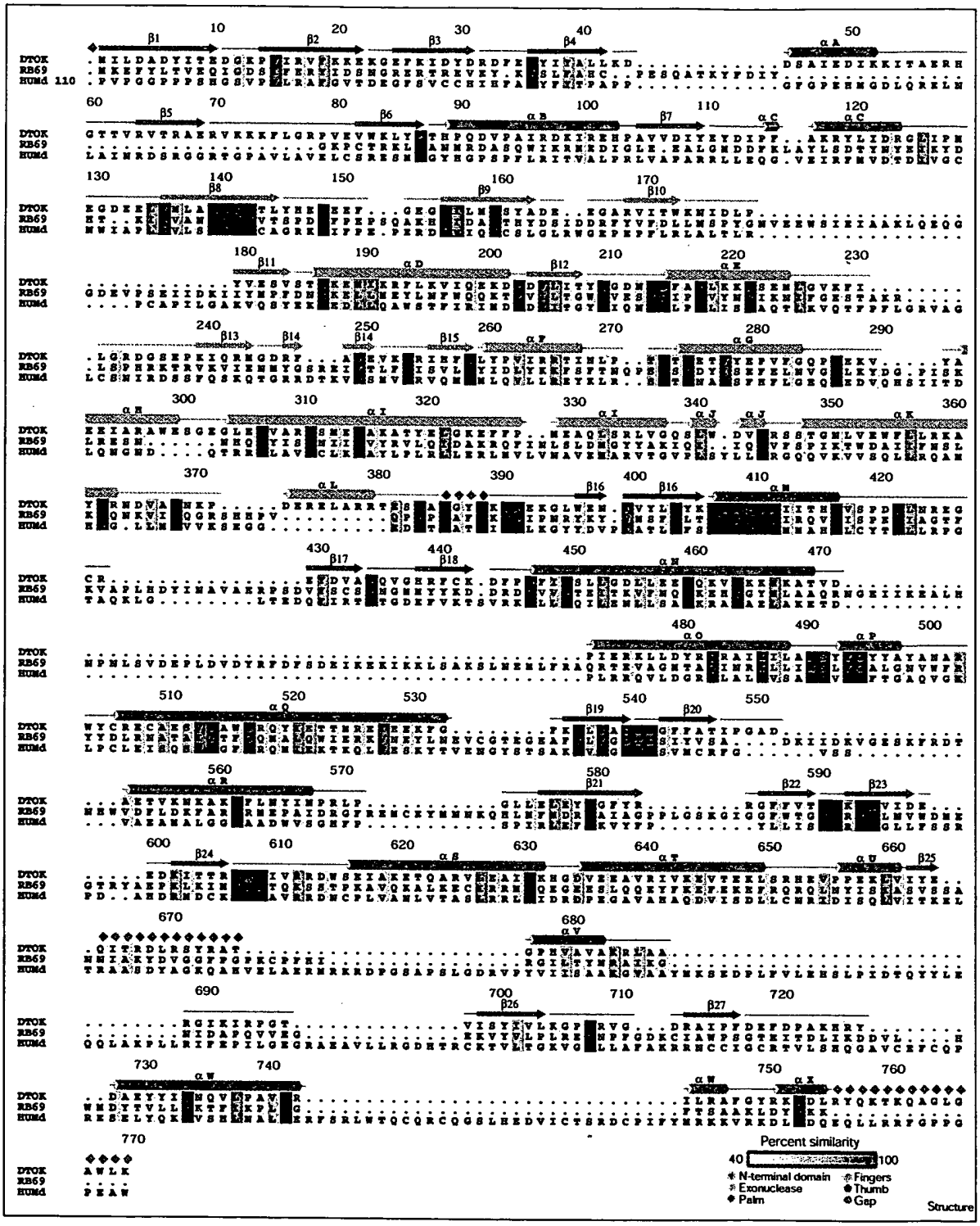

DTOK
RB69
HUMd
Percent similarity
40
100
N-terminal domain
Exonuclease
Palm
Fingers
Thumb
Gap
Structure

Figure 5

Structure-based sequence alignment for D. Tok Pol (DTOK), RB69 Pol (RB69) and human Pol δ (HUMd). The HUMd sequence begins at residue 110, as indicated by the number at the beginning of the sequence. The alignment is colored by sequence similarity (40%, white to 100%, green) calculated as described in Figure 7c. Shown here is a small subset of a larger set of sequences that were used to generate the alignment. The full sequence alignment is available at http://www.rockefeller.edu/Kuriyan. The respective secondary structural elements colored as in Figure 1 are represented by helices as cylinders, strands as arrows, and other as thin lines. Gray circles represent portions of the polypeptide chain that could not be modeled.

closely (Figure 3). Not surprisingly, the regions of highest sequence similarity are concentrated in and around the exonuclease and polymerase active sites (Figures 2c,5). Despite the low overall sequence identity, the individual subdomains in the two structures superimpose well (the rmsd in Cα positions in the fingers, thumb and palm subdomains is in the range of 0.8 to 1.5 Å). Moreover, the overall arrangement of domains and subdomains with respect to each other is preserved in the two polymerases, strengthening the proposal that Pol II DNA polymerases share a common architecture (Figure 3).

One difference between the overall structures of D. Tok Pol and RB69 Pol concerns the orientation of the exonuclease domain with respect to the rest of the structure. When the two polymerases are superimposed on their respective palm subdomains it is seen that the exonuclease domain of RB69 is rotated inwards by ~8°, burying the active site in a solvent-inaccessible configuration [7]. In contrast, the exonuclease domain in D. Tok Pol has its active site essentially exposed to solvent. It is possible that conformational changes between open and closed configurations of the exonuclease domain are a part of the functional cycle of the protein, particularly as the two different forms of D. Tok Pol differ in the orientation of the exonuclease domain (not shown).

One interesting difference between D. Tok Pol and RB69 Pol is that the former is a thermostable DNA polymerase whereas the latter is not. Unfortunately, attempts to identify features in the D. Tok Pol structure that might be correlated with thermostability is complicated by the very low sequence similarity between the two enzymes. One feature that does stand out, however, is the increased formation of arrays of ionic interactions on the surface of D. Tok Pol when compared to that of RB69 Pol (Figure 6). The formation of networks of ionic interactions has been noted to correlate with thermostability in other proteins [16,32,33].

Generally, D. Tok Pol subdomains tend to be more compact, with smaller helices and shorter loops than are found in RB69 Pol, a feature that may be another important source of thermostability. For example, the palm subdomain displays close structural conservation of elements near the catalytic aspartate residues. However, helix αR in D. Tok Pol is much shorter that its counterpart in RB69 Pol, and a small substructure in front of the palm subdomain is entirely missing in D. Tok Pol (Figures 3,5). Deletion of these elements is also seen in a representative set of archaebacterial DNA polymerases [13,14]. Likewise, the fingers subdomain is missing a large mass of from its tip in D. Tok Pol (Figures 3,5). However, the RB69 fingers extension most probably plays a T4 phage-specific role, as it is also missing from our alignments of archaebacterial DNA polymerases and eukaryotic polymerases δ (Figure 5).

The N-terminal domain resembles RNA-binding domains

The N-terminal domain of D. Tok Pol has no corresponding element in Pol I type polymerases. Analysis of the

Figure 6

Comparison of surface charges in D. Tok Pol and RB69 Pol. Accessible-surface representation of (a) D. Tok Pol and (b) RB69 Pol in the same orientation after superposition of their palm subdomains. Surface regions corresponding to the terminal oxygen atoms of aspartate and glutamate are colored red, whereas surface regions contributed by the sidechain nitrogen of lysines and arginines are colored blue. D. Tok Pol has a striking pairing of oppositely charged residues not seen in RB69 pol. A representation of D. Tok Pol as a worm is included for orientation.

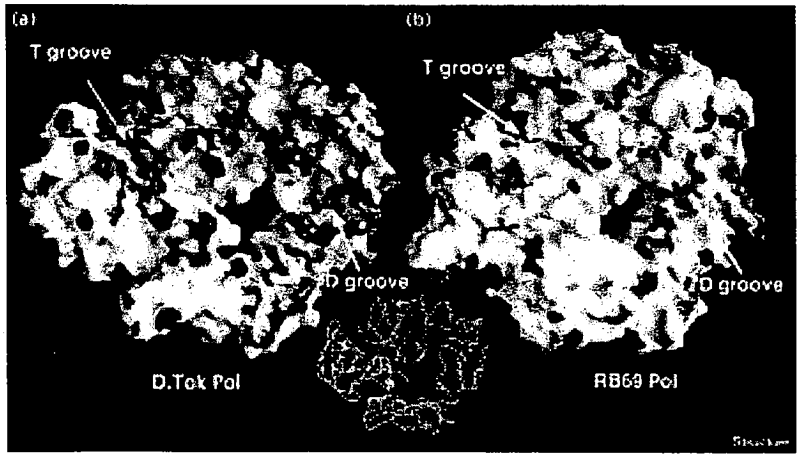
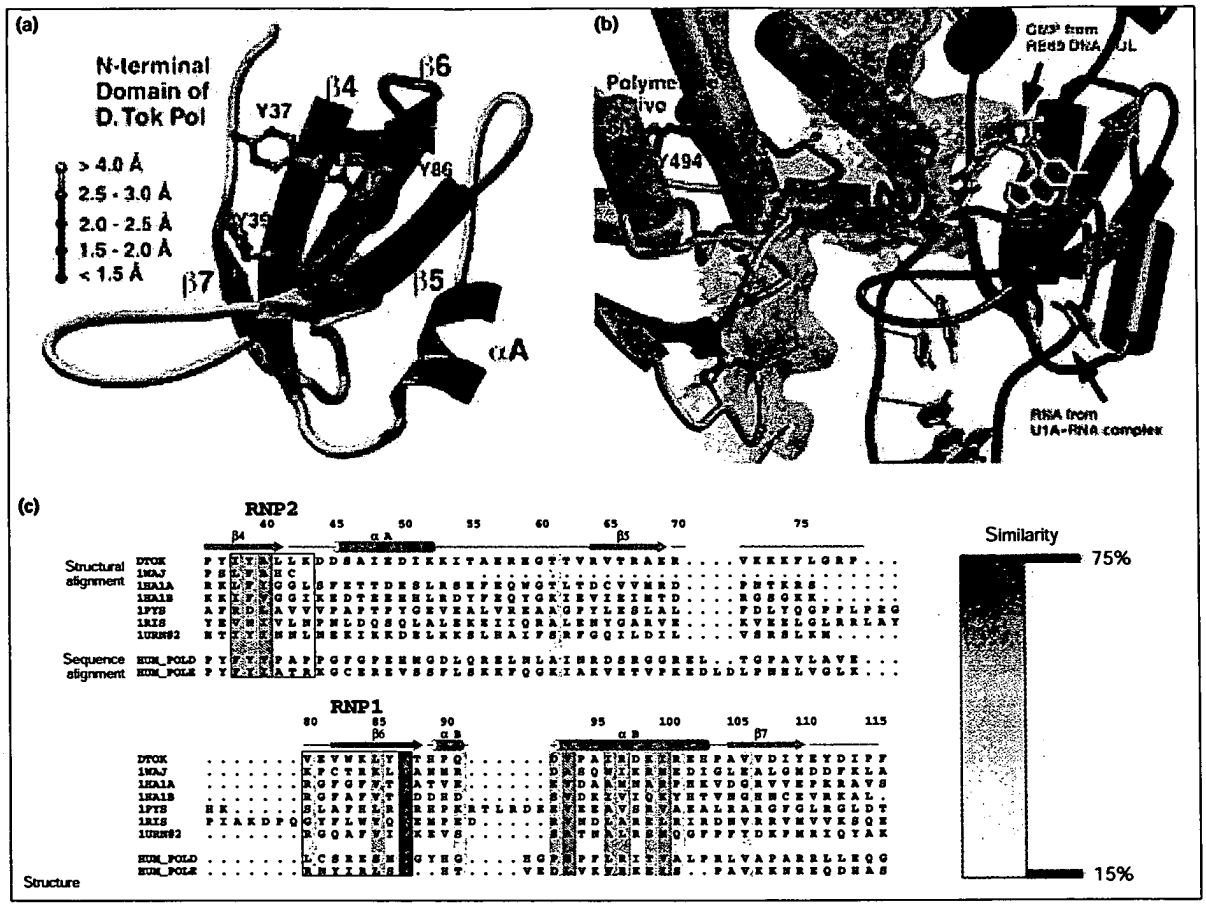


Figure 7



The N-terminal domain of D. Tok Pol. **(a)** Structural conservation of RNA-binding domains: structures of RNA-binding domains from 1HA1 (domains A and B), 1PYS, 1RIS, and 1URN (molecule 2) have been aligned by superimposing (LSQMAN, SUPERPOSE) D. Tok Pol residues 40–110, which represent four strands (β4, β5, β5, β6) and two helices (αA and αB). The N-terminal domain of D. Tok Pol is shown. A color gradient is used to depict the average rmsd in Cα positions for the family of superimposed structures, ranging from blue (1.0–1.5 Å) to white (> 4.0 Å). Certain aromatic residues in D. Tok Pol (white) are shown; these represent a potential RNA-binding surface. This view is rotated by approximately 180° from that in Figure 1. **(b)** An RNA stem-loop from the U1A–RNA complex (PDB code 1URN, [53]) modeled onto the N-terminal domain of D. Tok Pol. The model was generated by superimposing the U1A RBD onto the N-terminal domain of D. Tok Pol using the conserved structural elements. The RNA is drawn in blue with the sugar-phosphate backbone represented as a worm and the bases in ball and stick representation. A partial surface that represents the interface between the N-terminal domain of D. Tok Pol and the exonuclease domain is shown in gray. The location of the modeled RNA relative to the polymerase active site is depicted by marking the position of residue Y494. The location (derived after superposition) of the guanosine monophosphate (GMP) molecule bound to the 'incomplete' RBD of RB69 Pol, drawn in light green, nearly overlaps with the positions of the bases of the modeled RNA stem-loop. **(c)** Structural and primary sequence alignment of RNA-binding domains. Sequence alignment of the N-terminal domains from D. Tok Pol and RB69 Pol (incomplete domain) and the RBDs from 1HA1 (domains A and B), 1PYS, 1RIS, 1URN (molecule 2) superimposed as in Figure 7a. Alignments of the N-terminal domain of D. Tok Pol against DNA polymerase δ and ε were obtained using CLUSTALX [54], using its default parameters. The conserved primary sequence motifs RNP1 and RNP2 are boxed. The alignment is colored by sequence similarity (15%, white to 75%, green) calculated by averaging the similarity scores at each position of all possible pairs of sequences (DJ, unpublished software). Equivalence of nonidentical residues was established by use of the BLOSUM62 amino acid substitution matrix [55]. Secondary structural elements corresponding to the N-terminal domain of D. Tok Pol are represented (pink) with helices as cylinders, strands as arrows, and other as thin lines. Numbering of residues and naming of secondary structural elements is that of D. Tok Pol.

structure of this domain using DALI [34] (http://www.embl-ebi.ac.uk/dali/) revealed a previously unsuspected similarity to RBDs. RBDs are small modules (80–90 residues) found in RNA-binding proteins of prokaryotes, archaea, and eukaryotes (reviewed in [15]). These modules adopt a conserved βαββαβ architecture

and bind to single-stranded RNA. Two conserved sequence motifs, referred to as RNP1 (ribonucleoprotein 1) and RNP2, provide aromatic and charged residues that are important for RNA recognition [35] (Figure 7).

The N-terminal domain of D. Tok Pol can be superimposed closely onto the core secondary structural elements of RBDs from the U1A spliceosomal protein [35], ribosomal protein S6 [36], the heterogeneous ribonucleoprotein (hnRNP) proteins (two RBD domains) [37,38] and the anticodon-binding domain from *T. thermophilus* phenylalanyl-tRNA synthetase [39]. The rmsds in Cα positions for these superpositions are in the range of 0.5–2.0 Å (Figure 7a). Differences between the structures of the loops in the N-terminal domain of D. Tok Pol and those of the RNA-binding domains are within the range of structural variation seen in the various RNA-binding domains.

There is no evidence at present to suggest that the N-terminal domain of D. Tok Pol binds RNA. However, comparison with the structures of RNA complexes of RNA-binding domains shows that the N-terminal domain might in fact be a functional RNA-binding domain (Figure 7). In particular, three aromatic residues in the N-terminal domain (Tyr37, Tyr39 and Tyr86) could interact with RNA bases in a manner similar to that seen in crystal structures of RNA bound to RNA-binding domains [35] (Figure 7). Interestingly, these residues are located near the position of a guanosine triphosphate molecule that is found bound to the N-terminal domain of RB69 Pol [1] (Figure 7b). The DNA polymerases from bacteriophage T4 and its distant relative bacteriophage RB69 bind specifically to the ribsome-binding site of their own mRNA (messenger RNA), repressing its translation [40–42]. The N-terminal domains of T4 Pol and RB69 Pol are smaller than that of D. Tok Pol. In the RB69 Pol structure, the N-terminal domain seems to form an 'incomplete' RNA-binding domain (Figure 7c).

There is no significant overall sequence similarity between the N-terminal domain of D. Tok Pol and RNA-binding domains, which is why the presence of this fold was not recognized previously (Figure 7c). Comparison of the sequences of other archaebacterial DNA polymerases and human polymerases δ and ε suggests that a corresponding structural element is likely to be found in these polymerases as well (Figure 7c). The sequence alignment in this region is unambiguous for the archaebacterial DNA polymerases. For eukaryotic polymerases the alignment is less certain, but it seems to conserve the essential aromatic character of the RNP motifs (Figure 7c). Confirmation of the presence of these domains along with their ability to bind RNA, and their precise role in eukaryotic DNA synthesis awaits future structural and functional studies.

Biological implications

The structure of the DNA polymerase from the archaebacterium *Desulfurococcus* strain Tok, D. Tok Pol, reveals a strong similarity to the DNA polymerase from bacteriophage RB69. It also reveals the presence of an N-terminal domain that has structural similarity to RNA-binding domains from the U1A spliceosomal protein, ribosomal protein S6, the hnRNP proteins and the anticodon-binding domain from *T. thermophilus* phenylalanyl-tRNA synthetase. Although the structure in the immediate vicinity of the central catalytic region of the polymerase domain closely resembles that of Pol I type DNA polymerases, the overall architecture of D. Tok Pol and the placement of the exonuclease domain is strikingly different. The similarity between D. Tok Pol and RB69 Pol suggests that these two structures are representative of a common Pol II polymerase fold. Members of this family carry out chromosomal DNA replication in eukaryotes, including humans, and yet there is no structural information available for any eukaryotic member of this family. While this manuscript was being prepared, the structure of another archaebacterial DNA polymerase, that from the organism *Thermococcus gorgonarius* has been reported [56]. The D. Tok Pol structure reported here, along with the RB69 Pol structure and the structure of the *Thermococcus gorgonarius* DNA polymerase, should now make it possible to generate reliable structural models for eukaryotic DNA polymerases.

Materials and methods

Protein expression and purification

The D. Tok Pol bacterial expression vector and partial amino acid sequence were generous gifts of Life Technology Corporation. Convenient and reproducible protein expression was achieved by the cloning the D. Tok Pol gene into the Pet30 plasmid (Novagen). Determination of the amino acid sequence of the polymerase was completed using this construct. D. Tok Pol was purified by lysing biomass prepared from the above expression systems in a French pressure cell (Avestin). D. Tok Pol precipitated by incubation of the soluble fraction at 80°C for 30 min was further purified by ion-exchange (High-Q, Bio-Rad) and gel-filtration (Superdex-200, Pharmacia) chromatography. Purified protein was concentrated to 15 mg/ml by ultrafiltration (Millipore) in 40 mM TRIS-HCl, (pH = 7.4), 50 mM $(NH_4)_2SO_4$ for crystallization trials.

Crystallization, cryostabilization, and heavy-metal derivatization

Crystals of D. Tok Pol (maximum dimensions: 200 μm × 150 μm × 100 μm) were prepared from 100 mM TRIS-HCl (pH = 8.6), 10 mM $MgSO_4$, 200 mM $(NH_4)_2SO_4$, 20% (v/v) 2,4 methyl pentane diol (MPD), 11% (w/v) PEG4K, 10 mM dithiothreitol by vapor diffusion at 20°C. These crystals were cryostabilized in 100 mM TRIS-HCl (pH = 8.6), 10 mM $MgSO_4$, 200 mM Li_2SO_4, 20% v/v MPD, 13% w/v PEG4K for 30 min and when shock-cooled in freshly thawed liquid propane (−180°C), diffracted synchrotron wiggler radiation (A1 beamline, Cornell High Energy Synchrotron Source) to Bragg spacings of 2.4 Å. D. Tok Pol crystallized in space group $P2_12_12_1$ with cell parameters (Native I: a = 64.8 Å, b = 107.6 Å, c = 153.2 Å, α = 90°, β = 90°, γ = 90°). V_M calculations suggest that there is one molecule per asymmetric unit with high solvent content. Native data sets recorded under these conditions resulted in unacceptably high non-isomorphism between frozen samples. Substitution of PEG400 for MPD in the crystallization and stabilization media resolved this problem and allowed structure determination by multiple isomorphous replacement

(MIR) (Native II: a = 66.1 Å, b = 107.6 Å, c = 155.9 Å, α = 90°, β = 90°, γ = 90°). Heavy-metal derivatives were obtained by soaking Native II crystals in stabilizing solution containing 10 mM heavy-atom compound for 24 h.

Data collection and phase determination

X-ray diffraction data sets from a set of shock-cooled native and isomorphous heavy-atom derivatives were recorded at the Cornell High Energy Synchrotron Source (CHESS) beamline A1 (λ = 0.908). Data from Native I crystals (prepared with MPD) extended to a Bragg spacing of 2.4 Å with an R_{sym} = 4.6%. MIR analysis was conducted on Native II crystals (prepared with PEG400), which yielded data to beyond 2.6 Å. X-ray diffraction data were indexed, integrated and scaled using the HKL package [43].

The positions of heavy atoms were located manually by inspection of difference Patterson maps and checked by cross-phased difference Fourier maps. Experimental phases were calculated using these sites with the program SHARP [17]. In our hands, higher quality electron density maps were obtained by performing individual single isomorphous replacement (SIR) calculations in SHARP and combining the individual SIR phases sets using the program SIGMAA [19,44]. The experimental phases were improved and extended by solvent flipping and negative-density truncation as implemented in SOLOMON. This procedure (SHARP/SOLOMON) yielded electron-density maps of sufficient quality to allow the entire D. Tok Pol polypeptide to be traced unambiguously. This map was dramatically improved over a map calculated with MLPHARE/SOLOMON [19].

Model building and refinement

The initial molecular model was built into a 3.0 Å electron-density map using the interactive molecular graphics program O [45]. Model refinement was carried by conjugate gradient minimization, torsion-angle dynamics, and tightly constrained atomic temperature factor refinement in the program CNS [20]. Refinement against the 2.6 Å Native II data set was interspersed with manual rebuilding of the model against σ_A-weighted electron-density maps using $(2|F_o|-|F_c|)$ and $(|F_o|-|F_c|)$ coefficients calculated by averaging structure factors of ten models resulting from multiple torsion angle dynamics runs [46]. The original electron-density map remained a useful guide throughout the rebuilding process. The progress of the refinement was monitored by reductions in R_{free} (10% of the recorded reflections) [47]. Against the Native II data set, the model was refined to an R_{free} = 29.5% and $R_{working}$ = 24.2%. The refinement was continued against the 2.4 Å data Native I data set. A rigid-body search in CNS with the 2.6 Å model yielded a clear solution that was refined as above. The final model for Native I was refined to an R_{free} = 29.9% and $R_{working}$ = 25.3%, and the final model contains residues 1–756 with three disordered regions (386–389, 665–676, 757, 772). The Native II model contains 6030 non-solvent protein atoms, 4 sulfate ions, 2 magnesium ions, and 116 water molecules. The Native I model contains 5992 non-solvent protein atoms, 9 sulfate ions, 2 magnesium ions, and 106 water molecules. Model geometry was analyzed using the program PROCHECK [48]. Both models have no outliers in the Ramachandran plot, with over 80% of the residues in the most-favored region.

Figure preparation

Figures were composed in programs BOBSCRIPT v1.0 [49], GRASP v1.25 [50] and RIBBONS v3.00 [51], with renderings done in POVRAY v3.1e (http://www.povray.org). Figures 5 and 7c were composed using ALSCRIPT [52].

Accession numbers

Coordinates have been deposited with the Research Collaboratory for Structural Bioinformatics under the accession code 1QQC.

Acknowledgements

We thank Life Technologies, Inc. for providing the clone, partial sequence information and initial purification protocol of D. Tok Pol. DNA sequence determination was performed by the Protein/DNA Technology Center of the Rockefeller University. We thank S Jacques for technical assistance. L Berman, R Sweet, and the staff of the National Synchrotron Light Source and the Cornell High Energy Synchrotron Source are gratefully acknowledged for assistance with synchrotron data collection. Additional thanks go to J Bonanno, X Chen, E Conti, M Huse, S Soisson, T Schindler, A Boriack-Sjodin, SK Burley, and S Nair for help and advice. This work was supported in part by a grant from National Institutes of Health (GM45547-08; JK). Work at the Cornell High Energy Synchrotron Source (CHESS) is supported by the National Science Foundation under award DMR-9311772 and at the (MacCHESS) facility by award RR-01646 from the National Institutes of Health. YZ is supported by a Beckman Fellowship and a Burroughs Wellcome Fellowship.

References

1. Brathwaite, D.K. & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21**, 787-802.
2. Joyce, C.M. & Steitz, T.A. (1994). Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* **63**, 777-822.
3. Joyce, C.M. & Steitz, T.A. (1995). Polymerase structures and function: variations on a theme? *J. Bacteriol.* **177**, 6321-6329.
4. Brautigam, C.A. & Steitz, T.A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54-63.
5. Jaeger, J. & Pata, J. (1999). Getting a grip: polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **9**, 21-28.
6. Arnold, E., Ding, J., Hughes, S. & Hostomsky, Z. (1995). Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. *Curr. Opin. Struct. Biol.* **5**, 27-38.
7. Wang, J., Sattar, A.K., Wang, C.C., Karam, J.D., Konigsberg, W.H. & Steitz, T.A. (1997). Crystal structure of a Pol α family replication DNA polymerase from bacteriophage RB69. *Cell* **89**, 1087-1099.
8. Kelman, Z. & O'Donnell, M. (1994.) DNA replication: enzymology and mechanisms. *Curr. Opin. Genes Dev.* **4**, 185-195.
9. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. & Steitz, T.A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I Complexed with dTMP. *Nature* **313**, 762-766.
10. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.
11. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage replication complex at 2.2 Å resolution. *Nature* **391**, 251-257.
12. Lasken, R.S., Shuster, D.M. & Rashtchian, A. (1996). Archaebacterial DNA polymerases tightly bind uracil-containing DNA. *J. Biol. Chem.* **271**, 17692-17696.
13. Uemori, T., Ishino, Y., Toh, H., Asada, K. & Kato, I. (1993). Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.* **21**, 259-265.
14. Kong, H., Kucera, R.B. & Jack, W.E. (1993). Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus littoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J. Biol. Chem.* **268**, 1965-1975.
15. Varani, G. & Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 407-445.
16. Korolev, S., Nayal, M., Barnes, W.M., Di Cera, E. & Waksman, G. (1995). Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5 Å resolution: structural basis for thermostability. *Proc. Natl Acad. Sci. USA* **92**, 9264-9268.
17. De-La-Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. In *Macromolecular crystallography*.(Sweet, R.M. & Carter, C.W., eds), Vol. 276, pp. 472-493, Academic Press, New York.
18. Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30-42.
19. CCP4. (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
20. Brünger, A.T., *et al.*, & Warren, G.L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.

21. Jeruzalmi, D. & Steitz, T.A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.* **17**, 4101-4113.
22. Sousa, R., Rose, J. & Wang, B.C. (1994). The thumb's knuckle. flexibility in the thumb subdomain of T7 RNA polymerase is revealed by the structure of a chimeric T7/T3 RNA polymerase. *J. Mol. Biol.* **244**, 6-12.
23. Kim, Y., Eom, S.H., Wang, J., Lee, D.-S., Suh, S.W. & Steitz, T.A. (1995). Crystal Structure of *Thermus aquaticus* DNA Polymerase. *Nature* **376**, 612-616.
24. Jacobo-Molina, A., *et al.*, & Arnold, E. (1993). Crystal structure of human immunodeficiency virus type I reverse transcriptase complexed with double stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA* **90**, 6320-6324.
25. Goodrich, L.D., Lin, T.C., Spicer, E.K., Jones, C. & Konigsberg, W.H. (1997). Residues at the carboxy terminus of T4 DNA polymerase are important determinants for interaction with the polymerase accessory proteins. *Biochemistry* **36**, 10474-10481.
26. Berdis, A.J., Soumillion, P. & Benkovic, S.J. (1996). The carboxyl terminus of the bacteriophage T4 DNA polymerase is required for holoenzyme complex formation. *Proc. Natl Acad. Sci. USA* **93**, 12822-12827.
27. Steitz, T.A., Smerdon, S.J., Jaeger, J. & Joyce, C.M. (1994). A Unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* **266**, 2022-2025.
28. Kiefer, J.R., Mao, C., Braman, J.C. & Beese, L.S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304-307.
29. Brautigam, C.A. & Steitz, T.A. (1998). Structural principles for the inhibition of the 3'–5' exonuclease activity of *Escherichia coli* DNA polymerase I by phosphorothioates. *J. Mol. Biol.* **277**, 363-377.
30. Kiefer, J.R., *et al.*, & Beese, L.S. (1997). Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **5**, 95-108.
31. Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669-1675.
32. Yip, K.S., *et al.*, & Consalvi, V. (1995). The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147-1158.
33. Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J.N. (1995). 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure* **3**, 1295-1306.
34. Holm, L. & Sander, C. (1999). Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.* **27**, 244-247.
35. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.
36. Lindahl, M., *et al.*, & Muranova, T.A. (1994). Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **13**, 1249-1254.
37. Shamoo, Y., Krueger, U., Rice, L.M., Williams, K.R. & Steitz, T.A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat. Struct. Biol.* **4**, 215-222.
38. Xu, R.M., Jokhan, L., Cheng, X., Mayeda, A. & Krainer, A.R. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* **5**, 559-570.
39. Goldgur, Y., *et al.*, & Safro, M. (1997). The crystal structure of phenylalanyl-tRNA synthetase from thermus thermophilus complexed with cognate tRNAPhe. *Structure* **5**, 59-68.
40. Wang, C.C., Pavlov, A. & Karam, J.D. (1997). Evolution of RNA-binding specificity in T4 DNA polymerase. *J. Biol. Chem.* **272**, 17703-17710.
41. Tuerk, C., Eddy, S., Parma, D. & Gold, L. (1990). Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. *J. Mol. Biol.* **213**, 749-761.
42. Pavlov, A.R. & Karam, J.D. (1994). Binding specificity of T4 DNA polymerase to RNA. *J. Biol. Chem.* **269**, 12968-12972.
43. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **275**, 307-326.
44. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140-149.
45. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
46. Rice, L.M., Shamoo, Y. & Brünger, A.T. (1998). Phase improvement by multi-start simulated annealing refinement and structure factor averaging. *J. Appl. Crystallogr.* **31**, 798-805.
47. Brünger, A.T. (1992). The free R value: a novel statistical quality for assessing the accuracy of crystal structures. *Nature* **355**, 472-474.
48. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
49. Esnouf, R. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 133-138.
50. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and associations: insights from the interfacial and thermodynamic properties of the hydrocarbons. *Proteins* **11**, 281-296.
51. Carson, M. (1991). RIBBONS 2.0. *J. Appl. Crystallogr.* **24**, 958-961.
52. Barton, G.J. (1993). ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.
53. Nagai, K., Oubridge, C., Jessen, T.H., Li, J. & Evans, P.R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515-520.
54. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
55. Henikoff, S. & Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA* **89**, 10915-10919.
56. Hopfner, K.P., *et al.*, & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl Acad. Sci. USA* **96**, 3600-3605.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.

doi:10.1006/jmbi.2000.3728 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 299, 447–462

JMB

Crystal Structure of a Pol α Family DNA Polymerase from the Hyperthermophilic Archaeon *Thermococcus sp.* 9°N-7

A. Chapin Rodriguez†, Hee-Won Park†, Chen Mao and Lorena S. Beese*

Department of Biochemistry Duke University Medical Center, Durham NC 27710, USA

The 2.25 Å resolution crystal structure of a pol α family (family B) DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol) provides new insight into the mechanism of pol α family polymerases that include essentially all of the eukaryotic replicative and viral DNA polymerases. The structure is folded into NH_2-terminal, editing 3′-5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures (bacteriophage RB69 and the recently determined *Thermococcus gorgonarius*), but differ in their relative orientation and conformation.

The 9°N-7 polymerase domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparison of the orientations of the fingers and exonuclease (sub)domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the NH_2-terminal domain of 9°N-7 pol may be structurally related to an RNA-binding motif, which appears to be conserved among archaeal polymerases. The presence of such a putative RNA-binding domain suggests a mechanism for the observed autoregulation of bacteriophage T4 DNA polymerase synthesis by binding to its own mRNA. Furthermore, conservation of this domain could indicate that such regulation of pol expression may be a characteristic of archaea. Comparion of the 9°N-7 pol structure to its mesostable homolog from bacteriophage RB69 suggests that thermostability is achieved by shortening loops, forming two disulfide bridges, and increasing electrostatic interactions at subdomain interfaces.

Keywords: Archea; X-ray structure; replication; exonuclease; family B DNA polymerase

**Corresponding author*

Introduction

DNA polymerases catalyze the template-directed addition of nucleotides onto the 3′-OH group of the DNA primer terminus. These enzymes replicate DNA with the required accuracy essential for geno-

†Contributed equally to the manuscript.

Abbreviations used: pol, polymerase; *Tgo*, *Thermococcus gorgonarius*; ddNTP, dideoxyribonucles.

E-mail address of the corresponding author: lsb@biochem.duke.edu

0022-2836/00/020447–16 $35.00/0

mic stability, but generate sufficient mutations to stimulate and maintain evolution. Unlike Eucarya and Bacteria, relatively little is known about DNA replication in Archaea (Perler *et al.*, 1996), one of the three major evolutionary lineages of life (Woese *et al.*, 1990). Archea play a significant role in the biosphere, accounting for up to 30% of the biomass in certain Antarctic waters (De Long *et al.*, 1994), and exhibit much greater diversity than had originally been suspected (Barns *et al.*, 1996). Many characterized archaeal species are adapted to live in environments of extreme temperature, pressure, salinity, and/or pH such as hydrothermal vents, and hot springs (Rees & Adams, 1995).

Although archaeal cells share many morphological features with Bacteria, archaeal proteins involved in gene expression including DNA replication, transcription, and translation have been found to be similar to those from Eucarya (Edgell & Doolittle, 1997; Bult *et al.*, 1996). In particular, most of the archaeal DNA polymerases that have been sequenced belong to the α-like polymerase family (family B) that includes essentially all the eukaryotic replicative and viral DNA pols (Braithwaite & Ito, 1993; Edgell *et al.*, 1997).

Crystal structures exist for DNA pols from each of four families: pol I (family A), pol α (family B), pol β (family X) and reverse transcriptase (reviewed by Joyce & Steitz, 1994; Doublie *et al.*, 1999). Although pols from different families are structurally quite diverse, several common features have emerged. The pol domain from each resembles a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of *Escherichia coli* pol I (Klenow fragment) (Ollis *et al.*, 1985). All polymerases appear to share the same mechanism for nucleotidyl transfer involving two divalent metal ions (reviewed by Bautigam & Steitz, 1998). In addition, based on structures containing DNA and dNTP bound to pols from pol I, pol β, and reverse transcriptase families, a conformational change in the fingers subdomain from an open to a closed conformation is proposed to occur during the catalytic cycle (reviewed by Doublie *et al.*, 1999).

The pol α family polymerases are of medical importance as targets for development of antiviral and anticancer therapeutics. For example, human pol α is a target in the treatment of acute myelogenous leukemia and chronic lymphocytic leukemia (Keating *et al.*, 1982; Robertson & Plunkett, 1993) and a variety of nucleotide analogs with antitumor activity inhibit strand elongation by pol α (Huang & Plunkett, 1995; Gandhi & Plunkett, 1995). Furthermore, polymerases, particularly those that are thermostable, have a number of critical biotechnological applications ranging from PCR to cloning and DNA sequencing. Despite their biological, medical and biotechnological importance, the pol α class of polymerases has not been structurally as well characterized as other DNA polymerase families.

Here we report the 2.25 Å resolution crystal structure of a pol α family DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol). *Thermoccocus* sp. 9°N-7 was isolated from a hydrothermal vent at 9° N latitude off the East Pacific Rise (Southworth *et al.*, 1996). The structure is folded into NH_2-terminal, editing 3′-5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures (bacteriophage RB69 (Wang *et al.*, 1997) and the recently determined *Thermococcus gorgonarius* (*Tgo*) (Hopfner *et al.*, 1999), but differ in their relative orientation and conformation.

The pol domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the two other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparison of the orientations of the fingers and exonuclease domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit, and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the NH_2-terminal domain of 9°N-7 pol is structurally homologous to the βαββαβ RNA-binding motif with an exposed patch of aromatic amino acid residues. Bacteriophage T4 DNA pol, which is homologous to 9°N-7 pol, is known to bind its own mRNA and repress its own synthesis. The homology relationships to the RNA-binding motif suggest a structural basis for this regulatory mechanism. Furthermore, the conservation of this domain in other archaeal pols suggests that such autogenous regulation of pol expression may be general for archaea.

Results and Discussion

Crystal structure of *Thermococcus* sp. 9°N-7 pol

The structure of the full-length, 775-residue enzyme (bearing the double mutation D141A and D143A) was determined using the multiple isomorphous replacement method to a resolution of 2.25 Å. The current model has an *R*-factor of 23.9% (R_{free} = 30.8%) (Table 1). A Ramachandran plot of the model shows 86.8% of the residues in the most favored region and the remainder in additional allowed regions (12.4%) and generously allowed regions (0.8%). A total of 37 residues are not traced in the model and lie in regions of poorly defined electron density. The first of these gaps

Table 1. Crystallographic data collection and refinement statistics

Data	Native			Derivative						
	NAT-1[a]	NAT-2	NAT-3	THI-1[a]	THI-2	PtCl-1[a]	PtCl-2	PIP-1[a]	PIP-2	BAHg[a]
A. *Cell dimensions (Å)* ($P2_12_12_1$)										
a	96.1	94.8	95.3	96.7	96.2	95.8	96.3	97.0	95.4	96.1
b	101.1	98.2	98.7	100.7	99.6	100.8	99.9	100.5	98.6	100.3
c	112.2	112.2	112.6	112.5	113.2	111.9	112.4	111.6	112.3	111.1
Resolution (Å)	25-3.0	25-2.3	25-2.1	25-3.5	25-2.8	25-3.5	25-3.5	25-4.0	25-2.8	25-3.5
Highest-resolution bin (Å)	3.05-3.00	2.34-2.30	2.12-2.10	3.56-3.50	2.85-2.80	3.56-3.50	3.56-3.50	4.07-4.00	2.85-2.80	3.56-3.50
Completeness[b] (%)	99.0 (91.1)	89.2 (49.3)	94.2 (62.3)	99.0 (97.0)	86.2 (90.9)	98.8 (95.9)	97.3 (64.5)	99.6 (99.6)	96.6 (83.6)	97.9 (96.4)
Unique reflections	22,290	42,097	58,956	14,213	23,885	14,001	13,818	9592	25,789	13,749
Total reflections	164,954	312,366	652,420	177,293	575,987	177,578	119,768	130,698	255,691	220,642
R_{sym}[c]	8.2	5.1	5.5	9.1	10.1	7.3	11.7	11.3	14.6	8.0
R_{iso}[d]				17.7	12.9	19.5	37.1	24.2	15.3	26.0
Phasing power[e]				0.9	1.1	1.6	1.2	1.1	0.8 (to 4 Å)	1.0
B. *Refinement* ($F \geq 2\sigma F$)										
		rmsd from ideality								
Resolution limits (Å)	25-2.25									
Highest-resolution bin (Å)	2.35-2.25		In bond length (Å)		0.008					
Reflections	55-813		In bond angle (deg.)		1.260					
Completeness[b] (%)	93.5 (61.9)									
R-factor[f] (%)	23.9		Non-hydrogen atoms		5-918					
R_{free}[f]	30.8		Water molecules		109					

[a] Abbreviations: NAT, native; THI, sodium ethylmercurithiosalicylate; PtCl, K_2PtCl_4; PIP, di-μ-iodobis(ethylenediamine)diplatinum (II) nitrate; BAHg, Baker's mercurial.

[b] Numbers in parentheses correspond to completeness in the highest-resolution bin.

[c] $R_{sym} = (\Sigma|I - \langle I \rangle|)/\Sigma I$, where $\langle I \rangle$ refers to the average intensity I of multiple measurements of the same reflection.

[d] $R_{iso} = (\Sigma|I_{native} - I_{deriv}|)/\Sigma|I_{native}|$, where "native" and "deriv" refer to the native and derivative data.

[e] Phasing power = $\{\Sigma|F_H|^2/\Sigma|E|^2\}^{1/2}$, where $\Sigma|E|^2 = \Sigma(|F_{PH}|_{obs} - |F_{PH}|_{calc})^2$ and F_{PH} and F_H are the protein derivative and heavy atom structure factors, respectively.

[f] *R*-factor and $R_{free} = (\Sigma|F_{obs} - F_{calc}|)/|F_{obs}|$. R_{free} was calculated over 5% of the structure factor amplitudes not used in the refinement.

occurs at the bottom of the palm domain (residues 568-575), and the remainder are within the thumb region that is frequently observed to be partially disordered in apo polymerase structures, as is also the case here (e.g. Ollis *et al.*, 1985; Kiefer *et al.*, 1997). Although no disulfide bridges were included in the refinement, four Cys residues showed anomalous peaks in a difference Fourier map and side-chain distances and angles consistent with two disulfide bridges (Cys428:442, Cys506:509).

The structure of 9°N-7 pol reveals features common to all DNA pol structures as well as those that may be unique to archaeal pols. The overall shape of the enzyme can be described as a disc with a central hole that is folded into NH_2-terminal, 3′-5′ exonuclease, and polymerase domains (Figure 1(a) and (b)). Like all other pols of known structure, the pol domain resembles a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of *E. coli* pol I (Klenow fragment) (Ollis *et al.*, 1985). 9°N-7 pol is similar in structure to the pol α family polymerase from the mesostable bacteriophage RB69 (RB69 pol) (Wang *et al.*, 1997), although a number of these (sub)domains are shorter than in RB69 pol (Figure 1(c)). Nearly all these sequence length differences are attributable to loop segments that are fewer and shorter in the hyperthermostable 9°N-7. As was first observed in the RB69 pol structure (Wang *et al.*, 1997), the 3′-5′ exonuclease domain lies on the opposite side of the palm in comparison to pol I family polymerases. This domain arrangement is also seen in 9°N-7 pol and in *Tgo* pol (Hopfner *et al.*, 1999), indicating that this result is likely to be general for the pol α family. The structural similarity between 9°N-7 and RB69 pols is significant given the low sequence identity (<20%) in all but the active-site (palm) region, where sequence identity is 42% (Figure 2). Similar results hold for sequence alignments between 9°N-7 and human pol α.

NH_2-terminal domain

Many of the members of the pol α polymerase family, including archaeal pols, bacteriophage T4 and RB69 DNA pols, have an NH_2-terminal domain that is not observed in the pol I family. T4 pol is known to control its synthesis *in vivo* by a mechanism of autogenous regulation (Tuerk *et al.*, 1990). The mRNA-binding activity has been located to within the first 100 residues of the pol (Wang *et al.*, 1996), but the structure of a fragment comprising residues 1-388 of T4 pol failed to suggest a structural basis for RNA binding (Wang *et al.*, 1996). Here, we note that certain structural similarities between the homologous region in the 9°N-7 pol and the U1A RNA-binding protein may provide a rationale for RNA binding by T4 pol.

The NH_2-terminal domain of 9°N-7 pol can be considered as three modules based on compactness of folding (Figure 3(a)). The first module comprises residues 1-31, a three-stranded β-sheet that interacts extensively with the 3′-5′ exonuclease domain *via* predominantly electrostatic interactions. Residues 32-36 act as a flexible linker connecting the first module to the second (residues 37-123). The third module comprises residues 338-372.

The second module is folded into a βαββαβ motif, with two short β-strands, 5 and 6, inserted between the second and third elements. This motif occurs in a variety of proteins, and forms the basis for the most prevalent RNA binding motif, the RNA recognition motif (RRM). The RRM is present in the RNA-binding domains of hnRNP A1, spliceosomal protein U1A and U2B″, and the sex lethal protein (Burd & Dreyfuss, 1994). Although an alignment of the NH2-terminal domains of archaeal pols (Figure 3(b)), together with T4 and RB69 pols, shows that they lack the RNP1 and RNP2 sequence motifs that characterize the RRM (Burd & Dreyfuss, 1994), a number of highly conserved and invariant residues nevertheless emerges. Most of these residues fall in a cluster on the surface of the NH2-terminal domains of 9°N-7 and RB69 pols which therefore could mark the location of an RNA binding site atop the β-sheet platform on the face away from helix A (Figure 3(c)).

Both a sequence alignment (Figure 3(b)) and a structural comparison (Figure 3(c)) reveal that T4 and RB69 pols lack helix A and strand 7 of the βαββαβ motif, perhaps explaining why no suggestive structural homologies to RNA-binding folds could be identified (Wang *et al.*, 1996, 1997).

Experiments are needed to determine whether the NH_2-terminal domain of 9°N-7 pol binds RNA. Although the βαββαβ motif occurs in proteins that are not thought to interact with RNA (Burd & Dreyfuss, 1994), we find its presence in the NH2-terminal domain of 9°N-7 pol, in a region known to bind RNA in T4 pol (Wang *et al.*, 1996), to be highly suggestive of this. RNA-binding capability could hold for other archaeal pols as well, since sequence aligment of NH_2-terminal domain (Figure 3(b)) suggests that they share the βαββαβ motif.

We further speculate that just as T4 pol binds its mRNA to down-regulate its own synthesis, such autogenous regulation of pol expression might occur in archaea. Autogenous gene regulation is well documented in bacteria, and has at least one precedent in archaea. It has been identified in the synthesis of the MvaL1 ribosomal protein of *Methanococcus vanielii* (Hanner *et al.*, 1994), and postulated for a ribosomal gene cluster from the halophile *Halobacterium cutirubrum* (Shimmin & Dennis, 1989). It is interesting that there is no structural evidence that such regulation extends to eukaryotes, as human pol α shows no significant sequence homology to the NH_2-terminal sequences aligned in Figure 3(b).

3′-5′ Exonuclease domain

This domain is responsible for binding single-stranded DNA and excising mismatched bases in the elongated primer strand. The structure

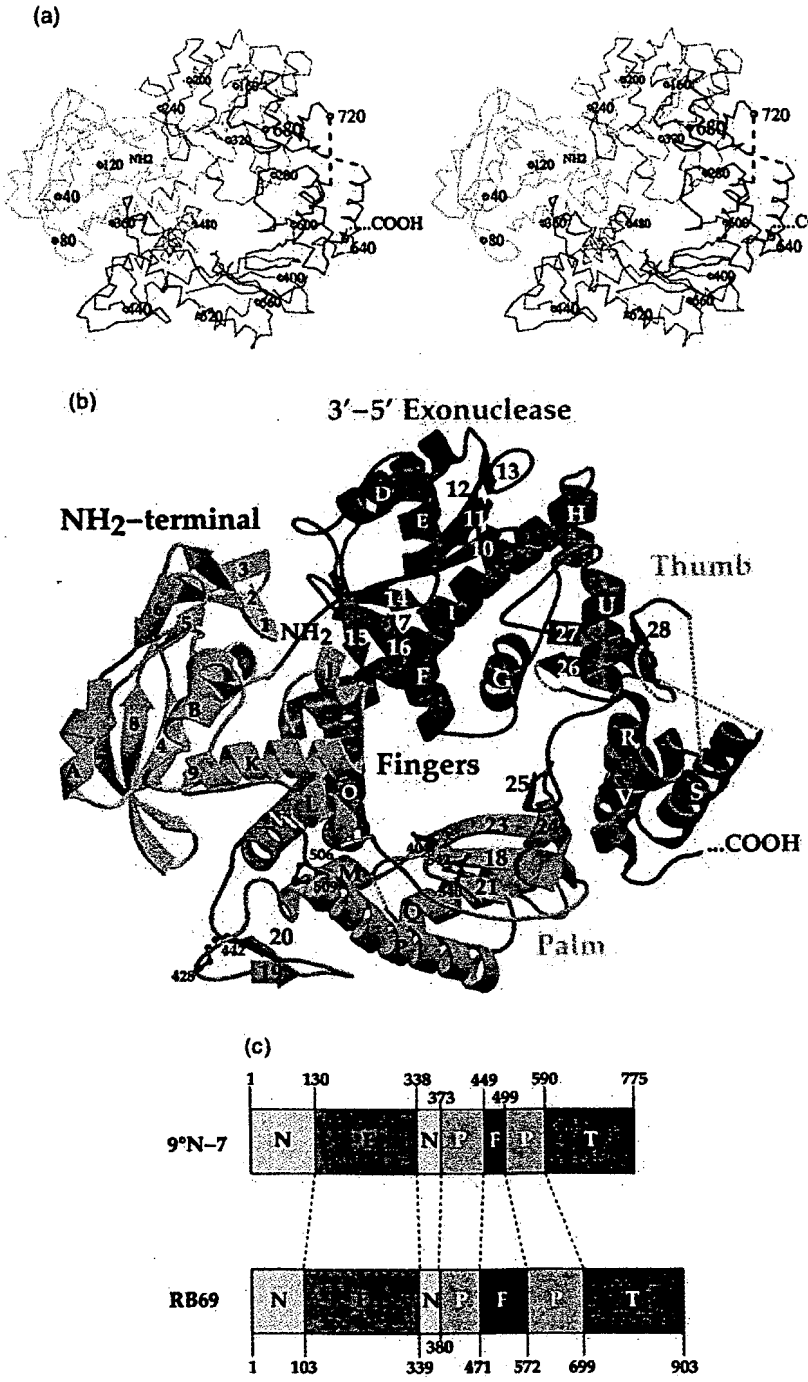


Figure 1. Structure of the *Thermococcus* sp. 9°N-7 DNA polymerase. The NH_2-terminal and 3′-5′ exonuclease domains are colored yellow and green, respectively. The polymerase domain is divided into palm (brown), thumb (red), and fingers (blue) subdomains. Three highly conserved carboxylate groups (D404, D540, D542) mark the polymerase active site. (a) Stereoview of the C^{α} trace. Every 40th C^{α} is numbered. Broken lines indicate disordered regions of the protein. (b) Ribbon diagram with secondary structure elements defined according to DSSP (Kabsch & Sander, 1983). NH_2-terminal domain: 1, 1-10; 2, 13-22; 3, 25-31; 4, 37-42; A, 48-51; 5, 55-58; 6, 61-64; 7, 67-75; 8, 78-86; B, 92-101; 9, 106-110; C, 116-123; J, 341-344; K, 349-363. 3′-5′ exonuclease domain: 10, 137-144; 11, 157-163; 12, 168-172; 13, 181-183; D, 187-201; 14, 205-208; E, 215-225; 15, 240-244; 16, 247-251; 17, 256-259; F, 260-266; G, 275-283; H, 292-300; I, 305-337. Polymerase domain: L, 374-379; 18, 397-404; M, 408-415; 19, 431-433; 20, 440-442; N, 448-468; O, 473-498; P, 507-532; 21, 535-539; 22, 543-547; Q, 553-567; 23, 578-590; 24, 593-598; 25, 603-606; R, 617-633; S, 636-651 (648-651 disordered); T, 657-660 (disordered); 26, 662-665; U, 677-688; 27, 698-703; 28, 714-716; V, 731-734; W, 742-746. (c) Schematic comparing the (sub)domains of *Thermococcus* sp. 9°N-7 7and bacteriophage RB69 DNA polymerases. The domain boundaries for 9°N-7 pol were determined based upon a structure-based sequence alignment with RB69 pol (Figure 2) as defined for the RB69 pol (Wang *et al.*, 1997).

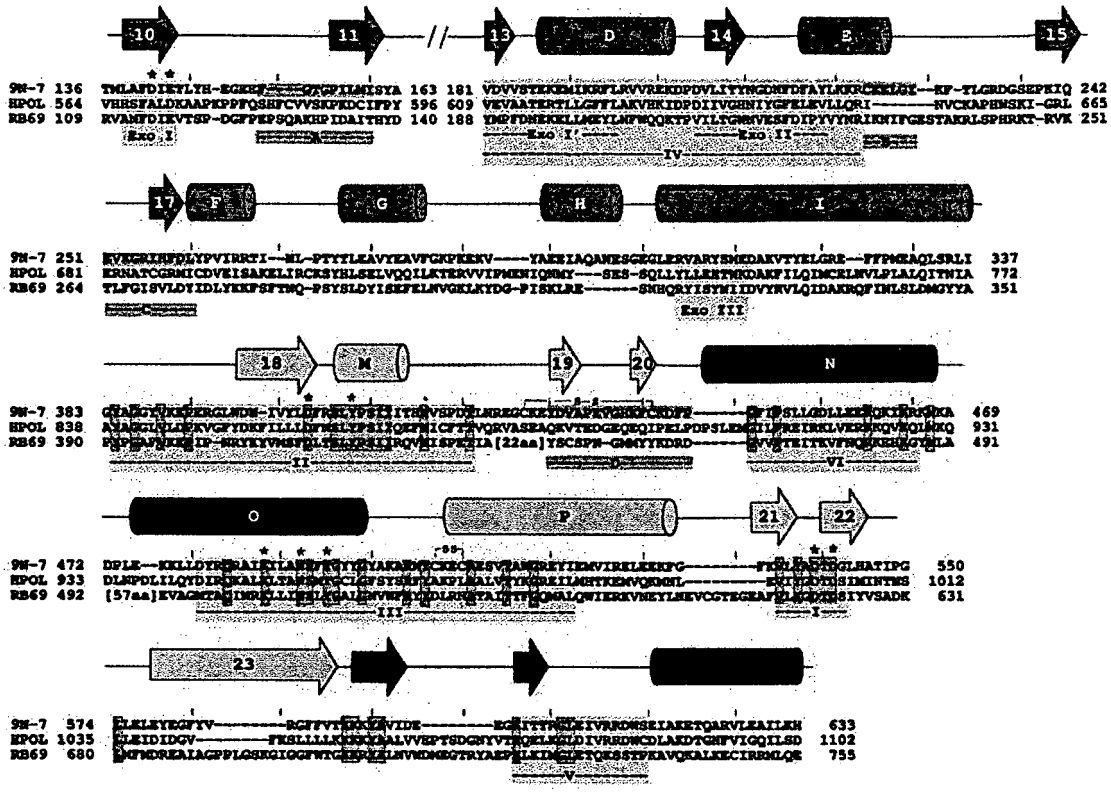


Figure 2. A three-way partial sequence alignment of *Thermococcus* sp. 9°N-7 pol (9N-7), RB69 pol (RB69), and human pol α (HPOL). Dashes indicate gaps in the alignment, and segments not aligned are represented as amino acid residue spans within brackets. Ticks mark every 10 spaces. The 9°N-7 and RB69 pol alignment is based upon the crystal structures. The HPOL and RB69 alignment is from Wang *et al.* (1997), except for a few short segments assigned based upon the three sequences shown here. Indicated below the sequences and boxed in yellow are consensus motifs in the exonuclease (Blanco *et al.*, 1992) and polymerase (Wong *et al.*, 1988) domains. The secondary structure elements in 9°N-7 pol, as defined by DSSP, are given above the sequences. The structural elements are colored according to the scheme described in legend to Figure 1. Shown in purple in the 9°N-7 pol sequence are the archaeal polymerase motifs described by Edgell *et al.* (1997). Residues within the polymerase domain that are invariant in the three sequences blue boxes; residues discussed in the section on dNTP binding, blue asterists. The two disulfide bridges in the palm (C428:C442, C506:C509) are shown schematically.

reported here is that of a mutant of 9°N-7 pol lacking detectable exonuclease activity which was engineered to prevent degradation of DNA substrates during subsequent co-crystallization experiments. This 9°N-7exo$^-$ pol was obtained by making two point mutations (D141A, E143A) in the Exo I (DxE) motif highly conserved among the 3′-5′ exonuclease domains of many DNA pols (Derbyshire *et al.*, 1995; Blanco *et al.*, 1992). In the Klenow fragment (KF) of *E. coli* DNA pol I, these residues (D355, E357) are responsible for binding the catalytic metals and for hydrogen-bonding with the 3′-OH of the terminal deoxynucleotide of the substrate DNA (Beese & Steitz, 1991).

Aside from loop segments that are shorter than those observed in RB69 pol (see below), the topology of the exonuclease domain in 9°N-7 pol is very similar to that of RB69 pol. The domains superimpose in the central β-sheet, containing the active site, with a root mean square deviation (rmsd) of 0.95 Å (35 C$^\alpha$ atoms). The metal-binding residues not mutated in 9°N-7exo$^-$ pol, D215 and D315, superimpose almost exactly on the corresponding RB69 pol residues (D222, D327).

It is now possible to assign a structural context to the four archaeal sequence motifs identified by Edgell *et al.* (1997). Three of the regions (A-C) lie within the exonuclease domain (Figure 2). Motif A forms part of the central β-sheet containing the active site; B, part of a solvent-exposed loop; and C, part of a five-stranded β-sheet nearly perpendicular to the central β-sheet. The fourth motif resides in the palm (see below).

Pol domain

This domain is responsible for the template-directed polymerization of dNTPs onto the growing primer strand of duplex DNA. Like other polymerases of known structure, the pol domain can be further divided into palm, fingers, and thumb subdomains. While the structure of the thumb of 9°N-7 and RB69 pols are highly similar, differences exist in the palm and fingers. Some of these differ-

ences correspond to features that appear unique to archaeal pols, while others support a hypothesis that a conformational change occurs in the fingers as part of the catalytic cycle.

Palm subdomain

The palm, which contains the active site for polymerization, shows a high degree of structural similarity to the palm subdomain of other DNA polymerases. It is as structurally similar to pol I family polymerases as to those of the pol α family. Its rms deviation from RB69 pol around the active site (blue region in Figure 4(b)) is 0.84 Å (26 C^{α} atoms). Together with the *Tgo* pol structure (Hopfner *et al.*, 1999), this structure confirms for archaea the conservation of a common catalytic core. A significant difference between the palm subdomains in 9°N-7 and RB69 pols are the two disulfide bridges present in 9°N-7 pol, one joining Cys428 and 442 and another joining Cys506 and 509 (Figure 4(b)). Both the shortened loops and at least one disulfide bridge appear common to archaeal pols (see above). Indeed, the region containing one of the Cys residue in a disulfide bridge (C442) corresponds to the highly conserved archaeal motif D (Edgell *et al.*, 1997; Figure 2). The *Tgo* pol structure shows the corresponding Cys residues to be "poised" for disulfide formation, but still in reduced form.

Until recently it was believed that all pols share a catalytic "triad" of carboxylate residues in the active site in the palm (Delarue *et al.*, 1990). Wang *et al.* (1997) since recognized that only two of the carboxylate residues are invariant. The invariant carboxylates in 9°N-7 pol are D404 and D542. The third member of the triad, present as D540 in 9°N-7 pol, is not essential: mutation at the corresponding residue (D1002N) in human pol α retains catalytic function (Copeland *et al.*, 1993). D540 in 9°N-7 pol may nevertheless be involved in binding the divalent metals required for catalysis. Mg^{2+} is normally the optimal metal for human pol α activity. The pol α D1002N mutant shows greater catalytic efficiency and fidelity with Mn^{2+} rather than Mg^{2+} (Copeland & Wang, 1993).

D540 in 9°N-7 pol interacts with the hydroxyl group of Y538 that is within hydrogen-bonding distance to D540. Substitution of this residue to Phe in human pol α (Y1000) causes only minor effects on catalysis but alters the pol metal affinity akin to the pol α D1002 mutation (Copeland & Wang, 1993). It seems likely that the hydroxyl moiety of Y538 in 9°N-7 pol helps to lock D540 in position for Mg^{2+}-specific binding. Consistent with this function is the strict conservation of Y538 among pol α family members (Braithwaite & Ito, 1993).

Fingers subdomain

The fingers subdomain of 9°N-7 differs in topology and relative conformation from RB69. The fingers of 9°N-7 pol are a simple helix-coil-helix, as in *Tgo* pol (Hopfner *et al.*, 1999), whereas in the fingers of RB69 pol, the coil region is expanded with more secondary structure elements (Figures 2 and 5). The shorter fingers of 9°N-7 pol are conserved among the archaeal pols aligned by Edgell *et al.* (1997). It is possible that the fingers of archaeal pols define a minimal functional unit.

Different positions of the fingers subdomain relative to the palm are observed in the 9°N-7 and RB69 pol structures (Figure 5(a)). The fingers of *Tgo* pol (Hopfner *et al.*, 1999) show a position intermediate between that in 9°N-7 and RB69 pols, when the palm subdomains of all three enzymes are aligned. It is interesting to note that the fingers subdomain of polymerases in the pol I family adopt different positions during the catalyic cycle (reviewed by Doublie *et al.*, 1999). An open position corresponds to that seen in the apoenzyme form (Ollis *et al.*, 1985; Kim *et al.*, 1995; Korolev *et al.*, 1995; Kiefer *et al.*, 1997) and the form bound to duplex DNA (Eom *et al.*, 1996; Kiefer *et al.*, 1998). A closed conformation has been observed in the ternary replication complexes of bacteriophage T7 pol (Doublie *et al.*, 1998), and Klentaq (Li *et al.*, 1998) with bound DNA and dNTP. An analogous conformational change has been observed in ternary complexes of human immunodeficiency virus reverse transcriptase (Huang *et al.*, 1998) and rat pol β (Pelletier *et al.*, 1994). In the closed conformation the fingers rotate towards the palm to form a binding pocket for dNTPs.

The differences in position of the fingers subdomain in the three pol α family crystal structures suggest that the fingers of pol α family pols move during catalysis, analogous to that observed for the other polymerase families. It is interesting to note that if this is the case, there must be a corresponding movement in the position of the 3′-5′ exonuclease domains not required in the other polymerase families as will be discussed below. If the position of the fingers in 9°N-7 pol more closely approximates a closed conformation, it is not clear why they would adopt a position previously observed only in ternary complexes with bound dNTP and DNA. The fingers of 9°N-7 pol may be stabilized in this conformation because of a salt-bridge between E578 in the palm and K487 on helix O of the fingers. These residues are highly conserved among archaeal pols (Edgell *et al.*, 1997) and both pol I and pols α families (Braithwaite & Ito, 1993). The corresponding salt-bridge does not form in polymerases of the pol I family because the fingers helix O lies too far from the palm. The fingers of *Tgo* pol, in fact, are rotated slightly away from the active site, relative to 9°N-7 pol, such that the E578:K487 salt-bridge cannot form. Another possible explanation for the difference in finger positions are the disulfide bridges present in 9°N-7 pol but absent in the *Tgo* pol structure and in pol I family structures. At least one of the disulfides (Cys428:442) in 9°N-7 pol could be directly involved in orienting the fingers relative to the palm (Hopfner *et al.*, 1999).

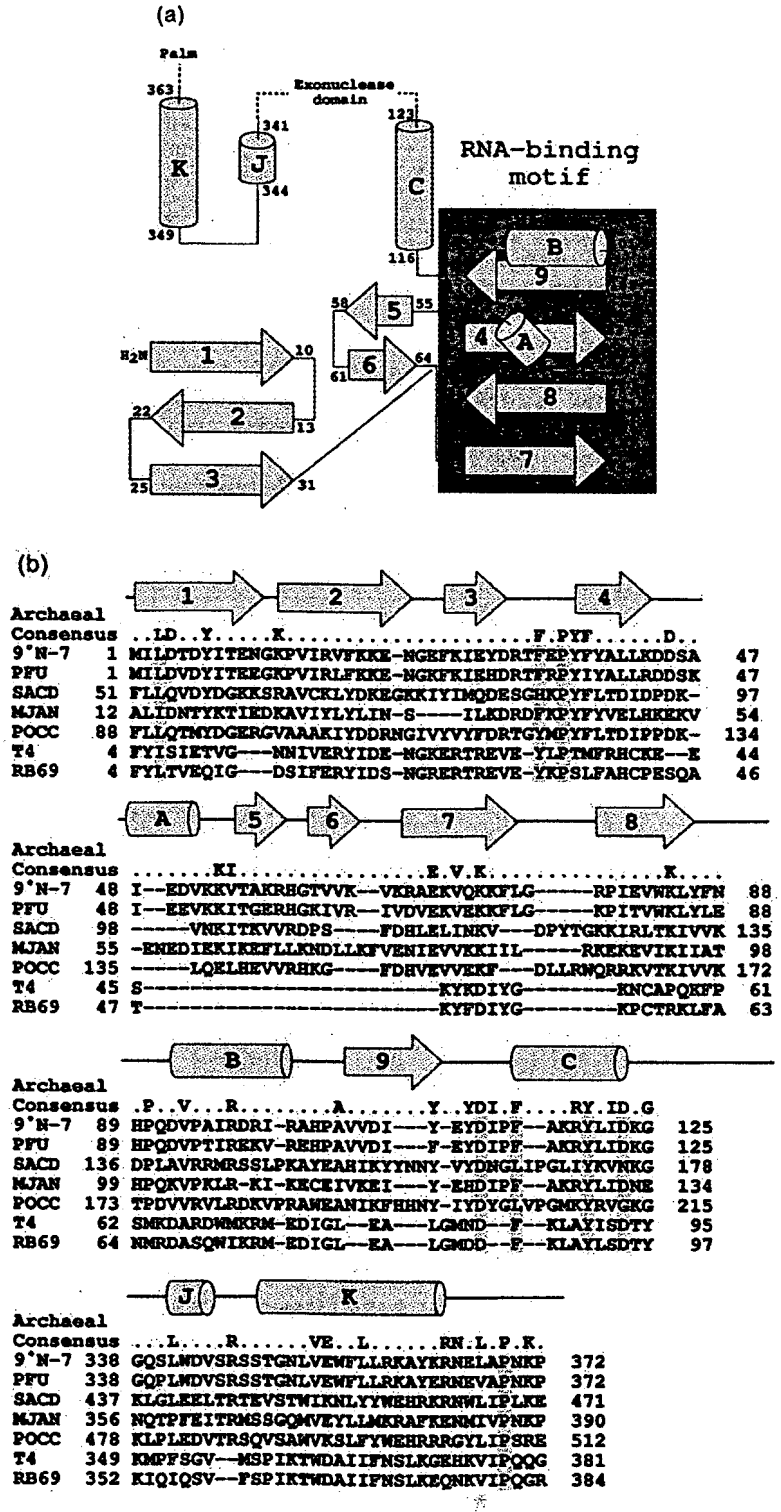


Figure 3 (*legend opposite*)

Model for DNA and dNTP binding

Based on the high degree of structural homology of the palm subdomains between 9°N-7 and pol I family pols, DNA and dNTP substrates from the bacteriophage T7 pol ternary complex (Doublie *et al.*, 1998) were modeled into the 9°N-7 pol active site. The model shown in Figure 6 provides further

C

9°N-7 RB69

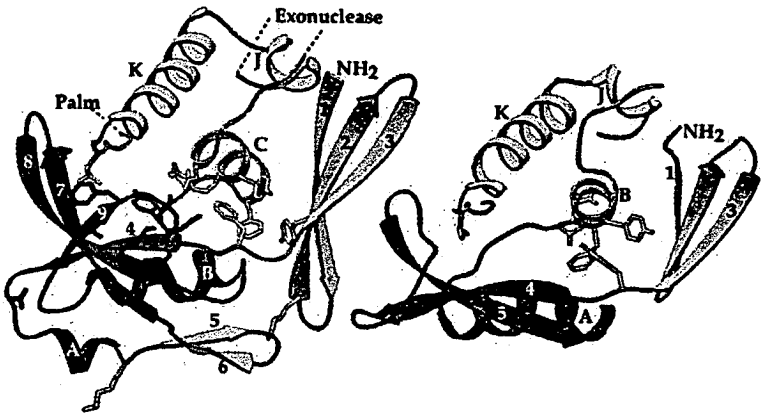


Figure 3. The RNA-binding motif in the NH_2-terminal domain of 9°N-7 pol. (a) Topology diagram of the complete NH_2-terminal domain (residues 1-129, 338-372). The RNA-binding motif βαββαβ known as the RNP recognition motif (Burd & Dreyfuss, 1994), is boxed. (b) Sequence alignment of the NH_2-terminal domains of 9°N-7 pol, RB69 pol, T4 pol, and archaeal pols. Alignment of 9°N-7 pol and T4/RB69 is based upon the crystal structures, and that of 9°N-7 pol and the other archaeal polymerases is based upon sequence alignment of 13 sequences among those considered by Edgell *et al.* (1997) (data not shown). The archaeal polymerase alignment was performed with the PILEUP algorithm in the GCG package (University of Wisconsin Genetic Computer Group). Secondary structure elements corresponding to 9°N-7 pol are given above the sequence. A consensus sequence was derived for the archaeal polymerases at those positions where at least 70% of the 13 sequences shared the same residue. Boxed in yellow are those residues conserved between the archaeal consensus and both bacteriophage (T4, RB69) sequences. Position 367 in 9°N-7 pol is starred (see the text for discussion). Abbreviations are as follows: PFU, *Pyrococcus furiosus*; SACD, *Sulfolobus acidocaldarius*; MJAN, *Methanococcus jannaschii*; POCC, *Pyrodictium occultum* B1. (c) Ribbons representation of the NH_2-terminal domain of 9°N-7 (left) and RB69 (right) pols. Least-squares C^{α} superposition was performed over the region of 9°N-7 pol including strand 4, part of strand 8, and helices B and C, and the domains were separated for side-by-side comparison. Shown in green is the βαββαβ RNA-binding motif. Charged and aromatic archaeal consensus residues are shown with green side-chains, and yellow side-chains correspond to the residues boxed in yellow in (b). The loop between β strands 7 and 8 in 9°N-7 pol corresponds to the conformationally variable loop 3 in the canonical RNP motif (Shamoo *et al.*, 1997).

evidence that the position of the fingers in 9°N-7 pol more closely approximates a closed conformation and their position in RB69 pol approximates an open conformation. This model of a ternary complex for a pol α family polymerase places the dNTP within hydrogen-bonding distance of residues on the fingers O helix that are highly conserved and known by mutagenesis to be functionally important. The corresponding residues on fingers helix P of the RB69 pol are farther away and cannot directly interact with dNTP.

The model places residues Y409 and Y494 near the deoxyribose moiety of the incoming dNTP. These residues appear to be functionally analogous to E480 and Y526 of T7 pol, which are responsible for discriminating between deoxy- and ribonucleotides (rNTPs). Y409 is invariant among the pol α family in the alignment by Braithwaite & Ito (1993) and nearly invariant (one exception) among archaeal pols aligned by Edgell *et al.* (1997). Mutation of the corresponding residue (Y412) to Val in an exonuclease-deficient *Thermococcus litoralis* (Vent) pol causes a 200-fold loss of discrimination against rNTPs. The aromatic ring appears to be the functionally important moiety, as mutating Y412 to Phe conserves wild-type discrimination levels (Gardner & Jack, 1999).

Y526 in T7 pol (F762 in Klenow fragment) has been dubbed the "ribose selectivity site" (Tabor & Richardson, 1995). A Phe residue at this position confers selectivity against incorporation of dideoxyribonucleotides (ddNTPs), whereas a Tyr residue in this position allows efficient incorporation of both nucleotide species. The presence of Tyr (Y494) in this position in 9°N-7 pol suggests the ability to incorporate dideoxynucleotides, as do Vent (Gardner & Jack, 1999) and human pol α (Copeland *et al.*, 1992). In fact, Tyr is invariant at this position among the archaeal pols aligned by Edgell *et al.* (1997), and highly conserved in the pol α family aligned by Braithwaite & Ito (1993).

The model of a ternary complex with dNTP and DNA places residues N491 and K487 in hydrogen-bonding distance from the triphosphate moiety of

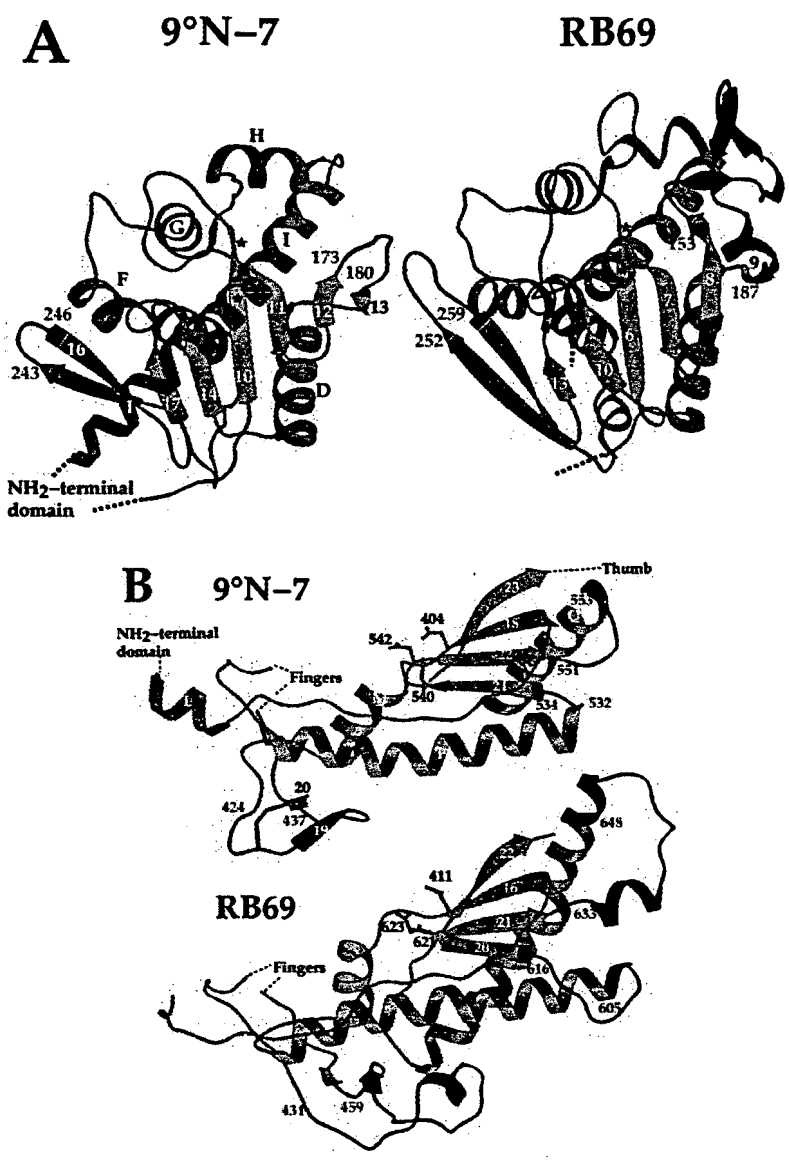


Figure 4. Comparisons of 9°N-7 and RB69 pols in different (sub)domains to indicate loop segments that are shorter in 9°N-7 pol. Least-squares C^{α} superposition was performed over the region in blue, and the domains were separated for side-by-side comparison. Loop regions are shown in magenta and their residue endpoints are labeled. (a) Comparison of the exonuclease domains. Indicated with purple asterisks are the active site carboxylates (mutated to Ala in the case of the 9°N-7exo$^-$ pol used in this study). (b) Comparison of the palm domains. The three active-site carboxylate groups are depicted with side-chains.

the incoming dNTP. Both of these residues are invariant in the pol α family (Braithwaite & Ito, 1993), and nearly invariant (one exception) among archaeal pols (Edgell *et al.*, 1997). Mutation of the corresponding residues (N494, K490) in Vent (exo-) pol severely decreases enzyme activity (Gardner & Jack, 1999).

Concerted domain movement

The difference in position of the fingers subdomain in 9°N-7 and RB69 pols is part of a larger conformational change involving the 3′-5′ exonuclease and NH_2-terminal domains. Comparing these two pol structures shows that in one of the pair, an essentially rigid-body rotation has occurred involving three of the five (sub)domains. This concerted movement affects both the position of the fingers relative to the pol active site (open *versus* closed conformation), as well as the position of the exonuclease active site relative to the pol active site. The 9°N-7 and RB69 pol structures may approximate different states along the reaction pathway corresponding to DNA synthesis and 3′-5′ exonucleatic proofreading activities.

When these two polymerases are aligned in the palm (the blue region in Figure 4(b)), the exonuclease and fingers are displaced between the proteins (Figure 5(a)). If the enzymes are aligned in the exonuclease domain (see Figure 4(a)), the fingers superimpose almost exactly (Figure 5(b)). Moving from a palm to an exonuclease-based alignment also brings the first module (residues 1-31) of the NH_2-terminal domains into identical positions (not shown). The joint motion of the first NH_2-terminal module and the exonuclease may reflect the need to maintain ionic networks at the interface. There are two five-membered ionic net-

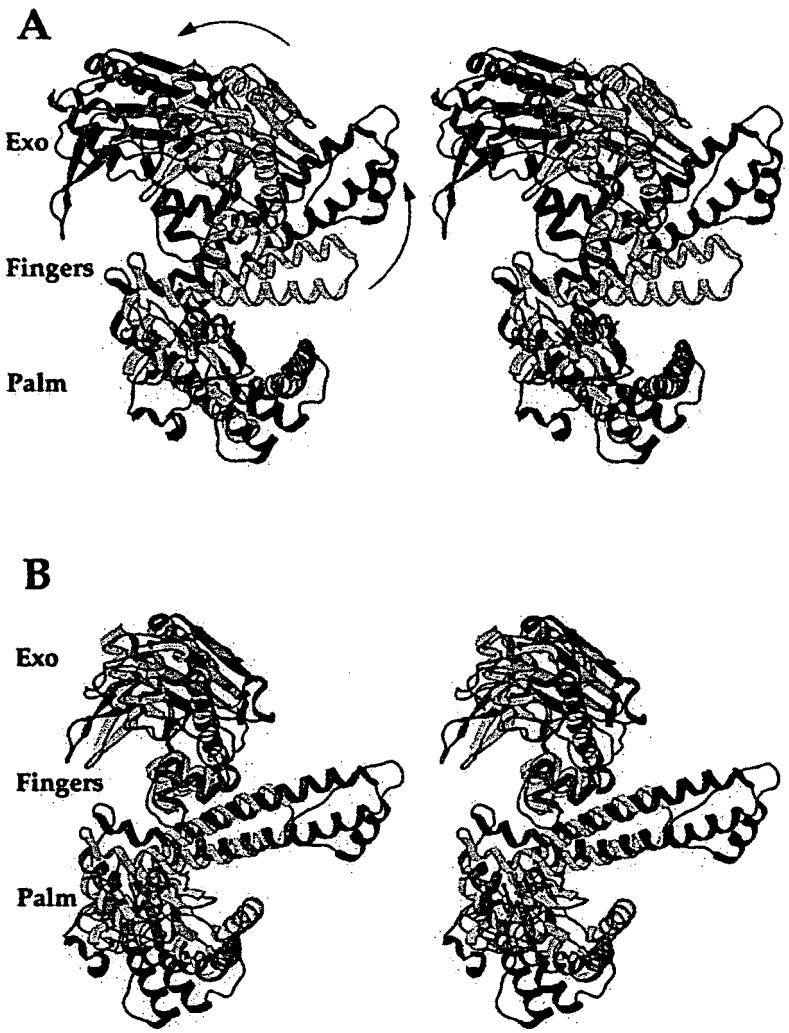


Figure 5. Least-squares C^{α} superpositions of 9°N-7 and RB69 pols in the (a) palm subdomain or (b) exonuclease domain. The 9°N-7 pol backbone is shown in yellow, and its active-site carboxylate groups in gold. The RB69 pol backbone is drawn in green, and its active-site residues in magenta. The central β-sheet of the exonuclease domain is light blue (9°N-7 pol) or dark blue (RB69 pol) to allow tracking of the domain motion. The precise regions used in the palm and exonuclease superpositions are shown in Figure 4. The NH_2-terminal domain has been omitted for clarity. Arrows in (a) indicate the direction of fingers and exonuclease movement when moving from (a) to (b).

works formed between the first module and exonuclease (Figure 7). In addition, a three-membered network is formed between the third NH_2-module (R346) and the exonuclease (Figure 7). This network is conserved among nearly all archaeal pols (Edgell *et al.*, 1997), but none is present in RB69 pol.

Comparison of the *Tgo* pol structure (Hopfner *et al.*, 1999) with that of 9°N-7 and RB69 pols using palm and exonuclease-based superpositions gives results similar to those in Figure 5, providing further support for the notion of a concerted domain movement.

A model was constructed for the RB69 pol (Wang *et al.*, 1997) showing how substrate DNA could shuttle between the pol and exonuclease active sites. When 9°N-7 and RB69 pols are aligned in the palm, the exonuclease active site in the former is tilted out and away from the pol active site, making it impossible for the DNA to shuttle. The exonuclease position in RB69, but not that in 9°N-7 pol, is therefore consistent with an editing conformation. It is interesting that this conformation also means that the fingers are not in position to bind dNTP (see above). Taken together, these considerations suggest that during the replication cycle of family B pols, there is concerted movement of the exonuclease, NH_2-terminal domain, and fingers relative to the catalytic region of the palm.

This concerted movement may be the structural basis for the functional coupling of polymerase and exonuclease domains, which is unique to the pol α family. In this family it is possible to generate site-directed mutations in one domain that exert an indirect, negative effect on the other (Reha-Krantz & Nonay, 1993; Abdus Sattar *et al.*, 1996). This contrasts with pol I pols like KF, where these activities are completely confined to their respective domains (Ollis *et al.*, 1985).

Molecular basis of thermostability

Thermococcus sp. 9°N-7 grows at temperatures of 88-90 °C, and its pol has a temperature optimum of 70-80 °C (Perler *et al.*, 1996). It has a half-life of 6.7

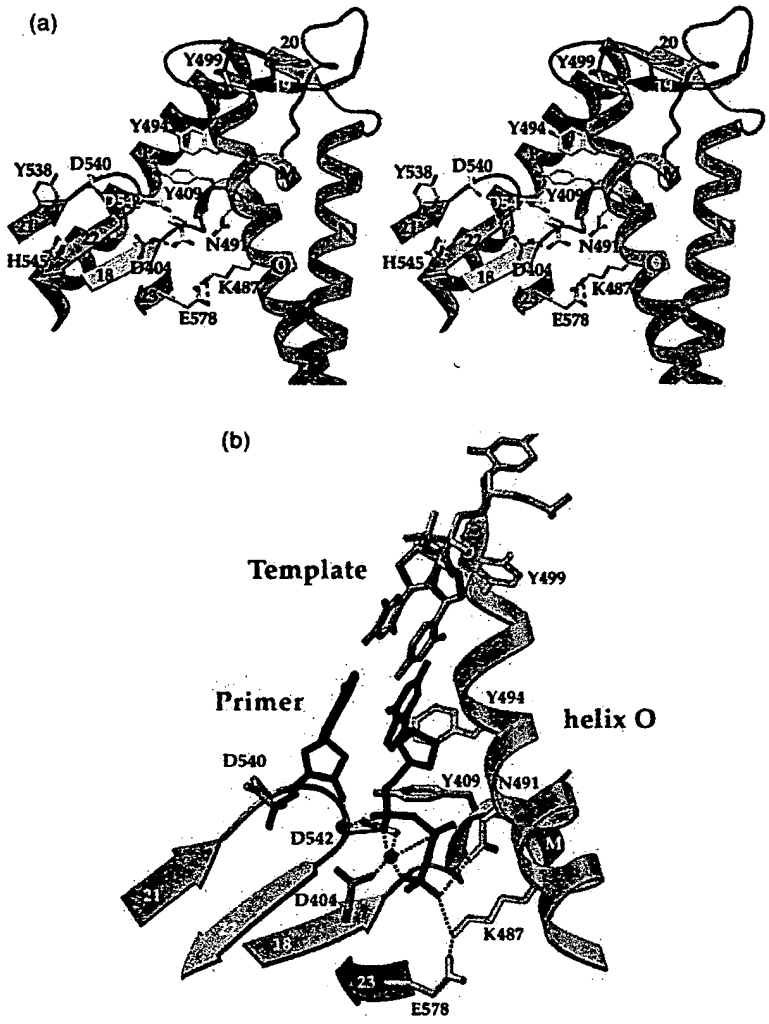


Figure 6. The active site of 9°N-7 pol and a modeled ternary complex. (a) Stereoview of the active site. Residues with indicated side-chains are discussed in the text. Hydrogen bonds are as broken lines, and the two disulfide bridges are shown in violet. K487 in this structure is involved in a salt-bridge with E578 of the palm. (b) Model of a ternary complex of 9°N-7 pol. For clarity only the incoming base and the first primer:template base-pair are shown. Hydrogen bonds are shown as broken lines, and metal ions are modeled as green spheres. The 9°N-7 pol and T7 pol ternary complex (Doublie *et al.*, 1998) were superimposed in the palm (0.55 Å rmsd for 13 C^{α} atoms). The rotamer conformation was adjusted for D542 and D404 in 9°N-7 pol, and the β turn including D542 was tilted downward, in a motion analogous to that observed between the apoenzyme and binary complex structures of *Bacillus stearothermophilus* pol (Kiefer *et al.*, 1997, 1998).

hours at 95 °C (R.B. Kucera, unpublished results), whereas *Thermus aquaticus* (*Taq*) DNA pol has a half-life of 1.6 hours at 95 °C (Kong *et al.*, 1993). The structure of 9°N-7 pol indicates a few key strategies for this hyperthermostability, some of which appear general to archaeal DNA pols.

A surprising feature of the 9°N-7 pol is that it contains two disulfide bridges (Figures 1(a) and 6(a)). The potential for the same bridges to form was also observed in *Tgo* pol (Hopfner *et al.*, 1999). Although not normally the case in Bacteria or Eucarya, an increasing number of cytosolic proteins with disulfide bridges are being discovered in the Archaea (DeDecker *et al.*, 1996; Singleton *et al.*, 1999). The stabilizing role of disulfide bridges has been well documented (Gokhale *et al.*, 1994; Cooper *et al.*, 1992). Introduction of disulfide bridges therefore appears to be a common strategy for archaeal protein stability.

Alignment of a large number of archaeal pols (Edgell *et al.*, 1997) suggests that having at least one of these disulfides is important for their thermostability. In fact, the two-stranded β-sheet containing C442 corresponds to sequence motif D in archaeal pols (Edgell *et al.*, 1997). Based on whether Cys is present in the corresponding positions, all the pols discussed by Edgell *et al.* (1997) are predicted to have at least one of the two disulfide bridges seen in 9°N-7 pol, with the exception of *M. voltae* and *S. shibatae* B3 pols. The mesostability of *M. voltae* pol may be partly caused by a lack of disulfide bridges. The *S. shibatae* B3 pol, like the *S. solfataricus* P2 B3 pol, is highly divergent in sequence from other archaeal pols, and it is unclear whether either of these functions *in vivo* (Edgell *et al.*, 1997).

An increased number of salt-bridges relative to mesostable homologs is often cited as a determinant of protein thermostability (DeDecker *et al.*, 1996; Korndorfer *et al.*, 1995; Chan *et al.*, 1995; Hennig *et al.*, 1995). The 9°N-7 pol shows a substantial increase in the fraction of charged residues participating in salt-bridges (47 %) compared with RB69 pol (39 %). These results are similar to a thermostability study of *Pyrococcus furiosus* glutamate dehydrogenase (Yip *et al.*, 1995). The authors of

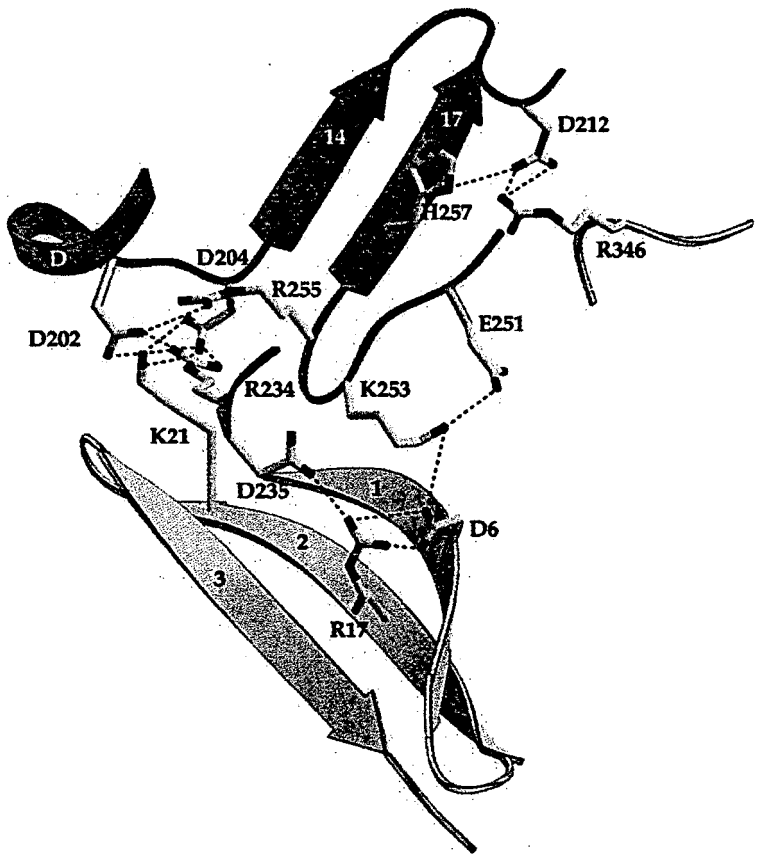


Figure 7. The extensive ionic networks at the interface of the NH_2-terminal and 3′-5′ exonuclease domains.

that study found a marked preference for Arg residues in the ionic interactions of the thermostable enzyme, but no such preference is evident here. The same fraction (48%) of Arg residues is used in ionic interactions in both 9°N-7 and RB69 pols, whereas a much higher proportion of Glu residues participate in salt-bridges in the 9°N-7 pol (53%) compared with RB69 pol (33%).

The number and distribution of salt-bridges within domains does not substantially differ between 9°N-7 and RB69 pols. At the interfaces between (sub)domains, however, the differences in ionic networks are striking. The proportion of ionic interactions at interfaces in the 9°N-7 pol (21%) is over twice that in RB69 pol (9%). The differences lie at the interface of the exonuclease domain with the NH_2-terminal domain (Figure 7), and at the interface of the exonuclease with the thumb, where a two and a three-member ionic network occur in 9°N-7 pol compared with none in RB69 pol (not shown).

Burial of the charged termini of proteins has been cited as another factor that can confer thermostability (Hennig *et al.*, 1995). The NH_2-terminal methione (M1) of 9°N-7 pol is stabilized by a hydrophobic cluster formed by L135, F327, I256, V205, and L341 while the corresponding residue of RB69 pol is completely exposed to solvent. The *B*-factor for the C^{α} of M1 in 9°N-7 pol is 26 Å^2, whereas for M1 in RB69 pol, it is 95 Å^2. While burial of the N terminus may be important for the thermostability of the 9°N-7 pol, the same does not hold for the C terminus. The last 25 residues are not visible in the electron density, similar to the case of RB69 pol. The solvent accessibility of the C terminus of these pols may reflect the need for this region to interact with a processivity accessory protein, which is known to be the case in the T4 replication complex (Berdis *et al.*, 1996).

Another common strategy for protein thermostability is to lower the solvent-accessible surface area of the protein and to increase the proportion of buried structure (Korndorfer *et al.*, 1995; Chan *et al.*, 1995). This translates into a more compact structural design. There are at least 12 examples of loop segments in RB69 pol that are much shorter or absent in 9°N-7 pol. Some of the more striking examples are shown in Figure 4. Alignment of 16 archaeal pols (Edgell *et al.*, 1997) indicates that they share practically all of these sequence "deletions". The *Tgo* pol structure also revealed shortened loop segments relative to RB69 pol (Hopfner *et al.*, 1999). Nevertheless, the overall ratio of solvent-accessible surface area to volume for both 9°N-7 and RB69 pols is the same (0.33). Thus, while lowering the surface area to volume ratio is a common strategy for thermostability, it is not the primary basis for the stability of 9°N-7 pol.

Materials and Methods

Purification, crystallization, and data collection

Thermococcus sp. 9°N-7 polymerase (wild-type and the D141A,D143A exonuclease-deficient mutant) was overexpressed and purified as described (Southworth *et al.*, 1996). Crystallization, cryoprotection, data collection and reduction of native crystals are described (Zhou *et al.*, 1998). Derivatives were prepared by soaking native crystals in stabilization solution (Zhou *et al.*, 1998) supplemented with 22.7 mM sodium ethylmercurithiosalicylate (thimerosal) for 11 days (thimerosal-1), 3.0 mM K_2PtCl_4 for one hour (PtCl-1), 1.5 mM di-μ-iodo-bis (ethylenediamine)diplatinum (II) nitrate (PIP) for one day (PIP-1), or 1.0 mM Baker's mercurial for 50 hours (BAHg). These crystals were stepped through stabilization solution containing 8% (five minutes), 16% (five minutes), and 30% sucrose (one to five hours). Additional derivatives were collected with the improved cryoprotection procedure reported (Zhou *et al.*, 1998) by soaking native crystals in 23.0 mM thimerosal for 8.3 days (thimerosal-2), 3.0 mM K_2PtCl_4 for seven days (PtCl-2), and 1.5 mM PIP for 35 hours (PIP-2).

Structure determination

The structure of the D141A,D143A mutant of 9°N-7 polymerase was determined by the method of multiple isomorphous replacement (MIR). A number of native and derivative crystals were used to solve the structure because of problems with non-isomorphism (Table 1). Three native datasets were collected from single crystals. NAT-1 was mounted in the liquid nitrogen stream directly from cryoprotectant, whereas NAT-2 and -3 were flash-frozen in liquid nitrogen prior to mounting. The crystals belong to space group $P2_12_12_1$ with unit cell dimensions of approximately $a = 96.1$ Å, $b = 101.1$ Å, $c = 112.2$ Å (for NAT-3). One molecule is present per asymmetric unit, giving a solvent content of approximately 60%.

A difference Patterson map of thimerosal-1 was calculated using the program FFT in the CCP4 suite (CCP4, 1994). One heavy-atom site for this derivative was identified with the program RSPS (Knight, 1989). This site was used to calculate initial phases for NAT-1 at 5 Å resolution using the program MLPHARE (Otwinowski, 1991). Difference Fourier synthesis with the initial phases revealed three sites for the PtCl-1 derivative. Two more sites for this derivative were discovered with the phases derived from both thimerosal-1 and PtCl-1. The correct handedness of the phasing information from these derivatives was determined using MLPHARE, and anomalous scattering data from the derivatives were included in the phase calculation. Three sites for the BAHg derivative and four sites for PIP-1 were obtained from difference Fourier maps calculated to 5 Å resolution. All of these heavy-atom sites were included in subsequent phase calculations with NAT-1. The high-resolution phasing limit was extended to 3.5 Å. Because of the high solvent content in the crystals, use of the solvent-flattening program DM (Cowtan, 1994), in combination with histogram matching, improved the phases substantially. A polyalanine model was built into the improved electron density map of NAT-1 with the program O (Jones & Kjeldgaard, 1993) and refined in the program X-PLOR (Brünger, 1992). Phase combination using the program SIGMAA (Read, 1986) further improved the map during building and refinement.

Identification of side-chain densities was possible only after collecting a higher-resolution native dataset (NAT-2), along with diffraction data for three more derivatives obtained under improved cryoprotection conditions (thimerosal-2, one site; PtCl-2, four sites; PIP-2, five sites). These derivatives were used to calculate MIR phases of NAT-2 to 3.0 Å resolution. Partial model phases of NAT-2 were calculated using the refined polyalanine model derived from NAT-1. Because of significant differences in unit cell dimensions between NAT-1 and 2, it was first necessary to subject NAT-2 to rigid-body refinement against NAT-1 in X-PLOR. Combination of the polyalanine model phases and MIR phases with SIGMAA improved the electron density map. Model building, refinement, and phase combination were reiterated until a complete polyalanine model could be built. In the final stage of refinement, NAT-3 was used to extend the resolution limit to 2.1 Å and water molecules were added.

Coordinate files and illustrations

The *Thermococcus* sp. 9°N-7 polymerase atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the accession code 1QHT. The RB69 coordinates used for comparisons in this manuscript are those of the orthorhombic crystal form (accession code 1WAJ). Figures were prepared within the IRIS Showcase program (Silicon Graphics, Inc.) entirely (1(b), 2, 3(a) and 3(b)) or with images imported from MOLSCRIPT (1(a)) (Priestle, 1991) or SETOR (3(c), 4-7) (Evans, 1993).

Acknowledgments

We thank New England Biolabs, Inc. for their collaboration on this project especially S. Kay Williams and Rebecca Kucera for protein purification and Francine Perler and William Jack for helpful discussions. We thank Airlie McCoy, David Owen, Daniela Stock, Jean-Luc Jestin, and Homme Hellinga for critical comments on the manuscript. We also thank Airlie McCoy, Jeff Taylor, and Sean Johnson for assistance in figure preparation. This work was supported by grants to L.S.B. from the ACS (NP-873A), the North Carolina Biotechnology Center, and the Searle Scholars Program.

References

Abdus, Sattar A. K., Lin, T. C., Jones, C. & Konigsberg, W. H. (1996). Functional consequences and exonuclease kinetic parameters of point mutations in bacteriophage T4 DNA polymerase. *Biochemistry*, **35**, 16621-16629.

Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily, and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA*, **93**, 9188-9193.

Beese, L. S. & Steitz, T. A. (1991). Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25-33.

Berdis, A. J., Soumillion, P. & Benkovic, S. J. (1996). The carboxyl terminus of the bacteriophage T4 DNA

polymerase is required for holoenzyme complex formation. *Proc. Natl Acad. Sci. USA*, **93**, 12822-12827.

Blanco, L., Bernad, A. & Salas, M. (1992). Evidence favouring the hypothesis of a conserved 3'-5' exonuclease active site in DNA-dependent DNA polymerases. *Gene*, **112**, 139-144.

Braithwaite, D. K. & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.* **21**, 787-802.

Brünger, A. T. (1992). *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D. *et al.* (1996). Complete genome sequence of the methanogenic archaeaon, *Methanococcus jannaschii. Science*, **273**, 1058-1073.

Brautigam, C. A. & Steitz, T. A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrates. *Curr. Opin. Struct. Biol.* **8**, 54-63.

Burd, C. G. & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615-621.

Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science*, **267**, 1463-1469.

Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.

Cooper, A., Eyles, S. J., Radford, S. E. & Dobson, C. M. (1992). Thermodynamic consequences of the removal of a disulphide from hen lysozyme. *J. Mol. Biol.* **225**, 939-943.

Copeland, W. C., Lam, N. K. & Wang, T. S.-F. (1993). Fidelity studies of the human DNA polymerase α: the most conserved region among α-like DNA polymerases is responsible for metal-induced infidelity in DNA synthesis. *J. Biol. Chem.* **268**, 11041-11049.

Copeland, W. C. & Wang, T. S.-F. (1993). Mutational analysis of the human DNA polymerase α: the most conserved region in α-like DNA polymerases is involved in metal-specific catalysis. *J. Biol. Chem.* **268**, 11028-11040.

Cowtan, K. (1994). "DM": an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM News. *Protein Crystallog.* **31**, 34-38.

DeDecker, B. S., O'Brian, R., Fleming, P. J., Geiger, J. H., Jackson, S. P. & Sigler, P. B. (1996). The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* **264**, 1072-1084.

Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* **3**, 461-467.

DeLong, E. F., Wu, K. Y., Prezelin, B. B. & Jovine, R. V. M. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature*, **371**, 695-697.

Derbyshire, V., Pinsonneault, J. K. & Joyce, C. M. (1995). Structure-function analysis of 3'-5' exonuclease of DNA polymerases. *Methods Enzymol.* **262**, 363-385.

Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, **391**, 251-258.

Doublie, S., Sawaya, M. R. & Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure*, **7**, R31-R35.

Edgell, D. R. & Doolittle, W. F. (1997). Archaea and the origin(s) of DNA replication proteins. *Cell*, **89**, 995-998.

Edgell, D. R., Klenk, H.-P. & Doolittle, W. F. (1997). Gene duplications in evolution of archaeal family B DNA polymerases. *J. Bacteriol.* **179**, 2632-2640.

Eom, S. H., Wang, J. & Steitz, T. A. (1996). Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature*, **382**, 278-281.

Evans, S. V. (1993). SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics*, **11**, 134-138.

Gandhi, V. & Plunkett, W. (1995). Cytotoxicity, metabolism, and mechanisms of action of 2',2'-difluorodeoxyguanosine in Chinese hamster ovary cells. *Cancer Res.* **55**, 1517-1524.

Gardner, A. F. & Jack, W. E. (1999). Determinants of nucleotide sugar recognition in an archeon DNA polymerase. *Nucl. Acids Res.* **27**, 2545-2553.

Gokhale, R. S., Agarwalla, S., Francis, V. S., Santi, D. V. & Balaram, P. (1994). Thermal stabilization of thymidylate synthase by engineering two disulfide bridges across the dimer interface. *J. Mol. Biol.* **235**, 89-94.

Hanner, M., Mayer, C., Kohrer, C., Golderer, G., Grobner, P. & Piendl, W. (1994). Autogenous translational regulation of the ribosomal MvaL1 operon in the archaebacterium *Methanococcus vanielli. J. Bacteriol.* **176**, 409-418.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). 2.0 Å structure of indole-3-glycerolphosphate synthase from the hyperthermophile *Sulfolobus solfataricus:* possible determinants of protein stability. *Structure*, **3**, 1295-1306.

Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius. Proc. Natl Acad. Sci. USA*, **96**, 3600-3605.

Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669-1675.

Huang, P. & Plunkett, W. (1995). Fludarabine- and gemcitabine-induced apoptosis: incorporation of analogs into DNA is a critical event. *Cancer Chemother. Pharmacol.* **36**, 181-188.

Jones, T. A. & Kjeldgaard, M. (1993). *O Version 5.9: The Manual*, Uppsala University, Uppsala, Sweden.

Joyce, C. M. & Steitz, T. A. (1994). Function and Structure relationships in DNA-polymerases. *Annu. Rev. Biochem.* **63**, 777-822.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Keating, M. J., McCredie, K. B., Bodey, G. P., Smith, T. L., Gehand, E. & Freireich, E. J. (1982). Improved prospects for long-term survival in adults with acute myelogenous leukemia. *J. Am. Med. Assoc.* **248**, 2481-2486.

Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. & Beese, L. S. (1997). Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure*, **5**, 95-108.

Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304-307.

Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, **376**, 612-616.

Kong, H., Kucera, R. B. & Jack, W. E. (1993). Characterization of a DNA polymerase from the hyperthermophile Archaea *Thermococcus litoralis*. *J. Biol. Chem.* **268**, 1965-1975.

Korndorfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). The crystal structure of hologlyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 511-521.

Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E. & Waksman, G. (1995). Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5 Å resolution. *Proc. Natl Acad. Sci. USA*, **92**, 9264-9268.

Li, Y., Korolev, S. & Waxman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7527.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, **313**, 762-766.

Otwinowski, Z. (1991). In *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), vol. 80, Science and Engineering Research Council, Warrington, UK.

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and dCTP. *Science*, **264**, 1930-1935.

Perler, F. B., Kumar, S. & Kong, H. (1996). Thermostable DNA polymerases. *Advan. Protein Chem.* **48**, 377-435.

Priestle, J. P. (1991). A program to produce both detailed and schematic drawings for protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A*, **42**, 140-149.

Rees, D. C. & Adams, M. W. W. (1995). Hyperthermophiles: taking the heat and loving it. *Structure*, **3**, 251-254.

Reha-Krantz, L. J. & Nonay, R. L. (1993). Genetic and biochemical studies of bacteriophage T4 DNA polymerase 3′-5′ exonuclease activity. *J. Biol. Chem.* **36**, 27100-27108.

Robertson, L. E. & Plunkett, W. (1993). High-dose cytosine arabinoside in chronic lymphocytic leukemia: a clinical and pharmacologic analysis. *Leuk. Lymphoma*, **10**, 43-48.

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nature Struct. Biol.* **4**, 215-222.

Shimmin, L. C. & Dennis, P. P. (1989). Characterization of the L11, L1, L10, and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. *EMBO J.* **8**, 1225-1235.

Singleton, M. R., Isupov, M. N. & Littlechild, J. A. (1999). X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Thermococcus litoralis*. *Structure*, **7**, 237-244.

Southworth, M. W., Kong, H., Kucera, R. B., Ware, J., Jannasch, H. W. & Perler, F. B. (1996). Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on *Thermococcus sp.* 9°N-7 and mutations affecting 3′-5′ exonuclease activity. *Proc. Natl Acad. Sci. USA*, **93**, 5281-5285.

Tabor, S. & Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl Acad. Sci. USA*, **92**, 6339-6343.

Tuerk, C., Eddy, S., Parma, D. & Gold, L. (1990). Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. *J. Mol. Biol.* **213**, 749-761.

Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H. & Steitz, T. A. (1996). Crystal structures of an NH_2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry*, **35**, 8110-8119.

Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. *Cell*, **89**, 1087-1099.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA*, **87**, 4576-4579.

Wong, S. W., Wahl, A. F., Yuan, P. M., Arai, N., Pearson, B. E., Arai, K., Korn, D., Hunkapiller, M. W. & Wang, T. S.-F. (1988). Human DNA polymerase α gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eucaryotic replicative DNA polymerases. *EMBO J.* **7**, 37-47.

Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*, **3**, 1147-1158.

Zhou, M., Mao, C., Rodriguez, A. C., Kiefer, J. R., Kucera, R. B. & Beese, L. S. (1998). Crystallization and preliminary diffraction analysis of a hyperthermostable DNA polymerase from a *Thermococcus* archaeon. *Acta Crystallog. sect. D*, **54**, 994-995.

Edited by D. Rees

(Received 15 June 1999; received in revised form 24 March 2000; accepted 24 March 2000)

Appendix I

We have purified and characterized the Family BDNA polymerase from the archaeon Methanococcus maripaludis, cloned from ATCC 43000. This polymerase has a 41% sequence identity and 63% sequence similarity with Vent DNA Polymerase when analyzed using NCBI Blast 2 and the default parameters.

We performed the titration assay described in Example 1 of the patent application, using the Mma, Vent (exo-), and 9°N (exo+) DNA Polymerases. Experimental details and data are given in the attached figure.

For each of the three polymerases, a comparison of lanes using dideoxyCTP (ddCTP) with those using equivalent concentrations of acycloCTP (acyCTP) reveals shorter products in lanes utilizing acyCTP. These shorter products result from more efficient insertion of the acyCTP terminator compared to incorporation of the ddCTP terminator. Thus, all three polymerases incorporated acyCTP more efficiently than ddCTP.

Figure Legend

The ability of acyNTPs and ddNTPs to act as chain terminators was tested using a titration assay of the type described in Example 1. Incorporation of ddCTP was compared to that of acyCTP, respectively, using Methanococcus maripaludis DNA polymerase, 9°N (exo+) DNA polymerase and Vent® (exo-) DNA polymerases.

Incorporation of ddCTP and acyCTP was assayed by mixing 8 µl of reaction cocktail (0.025 µM 5′ [FAM] end-labeled #1224-primed M13mp18, 62.5 mM NaCl, 12.5 mM Tris-HCl (pH 7.9 at 25°C), 12.5 mM $MgCl_2$, 1.25 mM

dithiothreitol, *Methanococcus maripaludis* DNA polymerase or 0.125 U/µl 9°N (exo+) DNA polymerase or 0.125 U/µl Vent® (exo-) DNA polymerase) with 2 µl of 5X nucleotide analog/nucleotide solution to yield the final ratios of analog:dNTP indicated in the figures. After incubating at 72°C for 20 minutes, the reactions were halted by the addition of 10 µl formamide. Samples were then heated at 72°C for 3 minutes and a 1 µl aliquot was loaded on a 4% polyacrylamide urea gel and detected by an ABI377 automated DNA sequencer.

ddCTP v. acyCTP incorporation by archaeal DNAPs